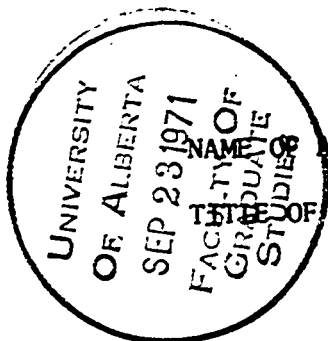


9641

NATIONAL LIBRARY
OTTAWABIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR..... ROBERT M. SANFORD
 TITLE OF THESIS..... THE RESPONSE OF HIGH
 SCHOOL STUDENTS TO POETIC
 LANGUAGE
 UNIVERSITY..... U. of ALBERTA
 DEGREE FOR WHICH THESIS WAS PRESENTED..... PH.D.
 YEAR THIS DEGREE GRANTED..... 1971

Permission is hereby granted to THE NATIONAL LIBRARY
 OF CANADA to microfilm this thesis and to lend or sell copies
 of the film.

The author reserves other publication rights, and
 neither the thesis nor extensive extracts from it may be
 printed or otherwise reproduced without the author's
 written permission.

(Signed)..... *R.M. Sanford*

PERMANENT ADDRESS:

625- South Drive
 Winnipeg 19, Man.

DATED..... *Sept 21* 1971

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

AN INVESTIGATION OF THE RESPONSE OF
HIGH SCHOOL STUDENTS TO POETIC LANGUAGE

by



ROBERT M. SANFORD

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

FALL, 1971

UNIVERSITY OF ALBERTA
FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "An Investigation of the Response of High School Students to Poetic Language" submitted by Robert M. Sanford in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

James B. Peck
Supervisor

W. M. Petridge

H. Kass

G. W. Balstrand

St. Armstrong
External Examiner

Date August 14, 1971

ABSTRACT

In order to investigate the response of high school students to poetic language an experimental poetry test using a "variant versions" technique was developed and administered to a sample of Grade X and XII students. Interviews based on test responses were then conducted with a representative sub-sample of students.

The Poetry Test was constructed according to four categories labeled Thought, Suggestion, Sound and Rhythm. An attempt was made by means of a number of pilot studies to develop two parallel forms, each consisting of 16 items, four in each subtest.

Item analysis was conducted to determine the difficulty and discriminating power of each item and to provide detailed information about item response patterns. Intercorrelations among the four subtests and between the Poetry Test results and scores on a standardized reading test were obtained to determine possible patterns of correspondence. Two-way analyses of variance were conducted to compare the influence of grade level, sex, and general reading ability on Poetry Test performance. A principal components analysis was also performed in an attempt to identify underlying factors in response to poetic language. The recorded interviews were then analyzed in order to supplement the statistical findings.

The results indicated that the investigatory procedure employed in the study was a useful means for examining student response to poetic language. While the two forms were not statistically equivalent, the Poetry Test was found adequately reliable and valid for its purposes, capable of providing a wide range of detailed information.

Item analysis showed test items generally too difficult for the Grade X students but suitable for Grade XII students. The mean item discriminating level was higher for the Grade XII than for the Grade X students.

The low positive correlations found among performances on each of the subtests and between Poetry Test and Reading Test performance indicated that the construct measured is a complex of partially discrete but inter-related skills and that the two tests measure markedly different abilities.

Poetry Test performance was found to be more closely related to grade level than to reading level (as measured by the Nelson-Denny Test). No significant connection was found between sex and Poetry Test performance.

Factor analysis findings indicated three factors which appeared to correspond to "everyday", "poetic", and "scientific" uses of language. These factor patterns bore little relationship to test categories, based on a traditional organization of poetic elements.

The interview results showed that the respondents were generally able to give clear reasons for choices within the test items. There was found to be a close relationship between the "correctness of approach" and the correctness of the answer. The students interviewed displayed certain common characteristics, such as ignorance of poetic terminology, but the high scorers tended to approach the language of a poem on several levels of meaning while the low scorers tended to operate on the discursive level only. The interview findings also revealed that superior test performance was accompanied by enjoyment of the poems, and a disposition to read poems intensively.

ACKNOWLEDGEMENTS

For their valued assistance the author wishes to thank Dr. J. B. Bell, his advisor; Dr. H. Kass, Dr. J. W. Bilsland, Dr. P. McFetridge, and Dr. Ted Aoki, his committee members; and the School District of North Vancouver, where the final tests were administered. Special thanks go to his wife Jocelyn, who kept the kids quiet .

TABLE OF CONTENTS

Chapter	Page
I. THE PROBLEM	1
BACKGROUND OF THE PROBLEM	1
STATEMENT OF THE PROBLEM.	7
DEFINITION OF TERMS	7
QUESTIONS AND HYPOTHESES.	12
DESIGN OF THE STUDY	15
DELIMITATIONS OF THE STUDY.	17
LIMITATIONS OF THE STUDY.	18
SIGNIFICANCE OF THE STUDY	18
PLAN OF THE STUDY	19
II. SURVEY OF RELATED LITERATURE.	21
INTRODUCTION.	21
THE NATURE OF POETIC LANGUAGE	22
General Characteristics	24
Organic unity	25
Fluidity of meaning	29
Multiplicity of meanings.	34
Poetic language and poetic diction.	37
Summary	39
Specific Structural Elements of Poetry.	40
Connotation	41
Imagery	43

Chapter	Page
Metaphor	46
Sound	50
THE TESTING OF LITERATURE	53
STUDIES OF RESPONSE TO LITERATURE	57
POETRY TEST AND STUDIES OF RESPONSE	
TO POETRY	60
SUMMARY	65
III. EXPERIMENTAL DESIGN	67
INTRODUCTION.	67
CONSTRUCTION OF THE POETRY TEST	67
General Rationale	67
Development of Categories	70
Selection of Poems.	74
Construction of Test Items.	75
Item Selection.	76
Validation by Judges.	78
Construction of Parallel Forms.	79
The Pilot Study	80
Final Revision of Test Forms.	81
THE INTERVIEWING PROCEDURE.	82
THE READING TEST.	84
SAMPLING PROCEDURE	85
DATA COLLECTION	87
STATISTICAL TREATMENT OF THE DATA	87
SUMMARY	88

Chapter	Page
IV. RESULTS AND DISCUSSION OF STATISTICAL PROCEDURES.	90
INTRODUCTION.	90
POETRY TEST CHARACTERISTICS	91
Reliability	91
Comparison of the Test Forms as a Whole.	97
Characteristics of the Four Subtests.	104
Summary of Characteristics of Test Forms and Subtests.	130
Characteristics of Individual Items	133
PERFORMANCE ON THE POETRY TEST BY VARIOUS GROUPS.	199
Grade Level Differences	199
Sex Differences	206
Reading Level Differences	206
Equivalence of the A and B Groups	212
Summary of Group Performances	214
FACTORS IN RESPONSE TO POETIC LANGUAGE.	215
Introduction.	215
Principal Components Analysis	219
The Varimax Rotation.	220
Factor Analysis Results	221
Interpretation of Factors in the Light of Poetic Theory.	237
Summary of Findings Relating to Factor Analysis	240
CHAPTER SUMMARY	243

Chapter	Page
V. INTERVIEW RESULTS	246
INTRODUCTION.	246
INTERVIEW RESULTS AND THE POETRY TEST	248
Test Validity	248
Reliability	257
Weaknesses in Specific Items.	260
Interview Results and Response to Poetic Language.	263
Common Characteristics in Student Response.	264
Comparison of High and Low Scores	270
Diagnostic Value of the Interviews.	280
CONCLUDING REMARKS AND SUMMARY.	282
VI. SUMMARY, CONCLUSIONS, AND IMPLICATIONS.	285
SUMMARY OF THE RESEARCH	285
SUMMARY OF THE MAIN FINDINGS.	286
CONCLUSIONS RELATING TO THE INVESTIGATORY PROCEDURE	289
CONCLUSIONS RELATING TO STUDENT RESPONSE TO POETIC LANGUAGE	291
IMPLICATIONS FOR THE TEACHING OF POETRY . . .	294
SUGGESTIONS FOR FURTHER RESEARCH.	299
BIBLIOGRAPHY	303
APPENDIX A. THE POETRY TEST	313
APPENDIX B. SOURCES OF THE POEMS.	330
APPENDIX C. THE INTERVIEW SCHEDULE.	334

	Page
APPENDIX D. SAMPLE INTERVIEW TYPESCRIPTS.	337
APPENDIX E. THE JUDGES.	357
APPENDIX F. DETAILED TABLES	359

LIST OF TABLES

Table		Page
I	K-R 20 Reliability Coefficients for Forms A and B.	93
II	Comparison of Mean Item Data Figures for Forms A and B.	95
III	K-R 20 Coefficients Estimated for a Test of 32 Items	97
IV	Means and Standard Deviations for Forms A and B.	98
V	Significance of Difference of Form A Means and Form B Means.	98
VI	Comparison of Mean Difficulty Indices and Biserial Correlation Coefficients for Forms A and B.	100
VII	Mean Biserial Correlation Coefficients Using Subtest Totals.	101
VIII	Comparison of Intercorrelations Between Total Scores and Subtest Scores for Forms A and B.	102
IX	Correlation Between Poetry Test Scores and Reading Test Scores	104
X	Grade X - Grade XII Gains on the Poetry Test and the Reading Test	105
XI	Comparison of Means and Standard Deviations of Subtest Scores	106
XII	Item Difficulty Index Range Within Each Subtest of Forms A and B.	108
XIII	Rank Orders of Mean Subtest Scores by Test Forms and Grade Groups	110
XIV	Per Cent Increases of Grade XII Over Grade X for the Four Subtests	111

Table		Page
XV	Item Data Means for the Four Subtests for Forms A and B	114
XVI	Intercorrelations Among Subtest Scores for Forms A and B	117
XVII	Mean Intercorrelations Among the Poetry Subtests.	120
XVIII	Intercorrelations of Poetry Subtest Scores and Reading Test Scores	123
XIX	Significant Intercorrelations (Poetry Subtests and Reading Subtests)	124
XX	Significant Intercorrelations (Reading Subtests and Poetry Test).	127
XXI	Intercorrelations Between Reading Scores and Subtest One and Four Scores . .	128
XXII	Complete Item Data for Form A.	134
XXIII	Complete Item Data for Form B.	135
XXIV	Range of Biserial Correlation Coefficients Within each Subtest	137
XXV	Distribution of Difficulty Indices and Biserial Correlation Coefficients. . .	137
XXVI	Comparison of Pilot Study Item Data with Final Data	139
XXVII	Detailed Item Data for "The Rich Man". . . .	144
XXVIII	Detailed Item Data for "Portrait VIII" . . .	151
XXIX	Detailed Item Data for "The Pulley".	156
XXX	Detailed Item Data for--from "Tintern Abbey".	160
XXXI	Detailed Item Data for--from "Paid on Both Sides"	165
XXXII	Detailed Item Data for "Bereft".	169
XXXIII	Detailed Item Data for "Restaurant".	175

Table		Page
XXXIV	Detailed Item Data for "Central Park Tourney".	179
XXXV	Detailed Item Data for--from <u>Julius Caesar</u>	185
XXXVI	Detailed Item Data for "Study Hall".	190
XXXVII	Two-Way Analysis of Variance: Test Form by Grade Level.	200
XXXVIII	Two-Way Analysis of Variance: Sex by Grade Level	202
XXXIX	Two-Way Analysis of Variance: Reading Level by Grade Level	203
XL	Means and Standard Deviations for Poetry Test by Grade Level	205
XLI	Reading Test Means for Students Taking Form A and Form B	213
XLII	Factor Analysis Results for Form A Pearson Intercorrelations.	223
XLIII	Factor Analysis Results for Form B Pearson Intercorrelations.	227
XLIV	Description of Factors A, B, and C	229
XLV	Items with High Varimax Coefficients (Pearson)	230
XLVI	Difficulty Level Change for Items with High Factor Loadings.	241
XLVII	Recall, Correctness, and Dimension of Test Responses as Indicated by the Follow-up Interviews: Detailed Figures.	249
XLVIII	Comparison of Correctness of Response with Correctness of Dimension for Poetry Test Totals and Subtotals	252
XLIX	Comparison of Correctness of Response with Correctness of Dimension for Items Loading High on Factors A, B, and C.	256

L	Comparison of Correctness of Response and Correctness of Dimension for High and Low Scorers	256
LI	Relation of Accuracy of Recall to Correctness of Answers and Corrections of Dimension	258
LII	Summary Percentage Figures for Interview Results Based on Table XLVII.	266
LIII	Capsule Characterization of High and Low Scorers	272
LIV	Comparison Between Low and High Scorers' Identification of "Correct" Dimension for Items in the Four Subtests	277
LV	Form A Item Data Obtained by Using Subtest Total Scores	351
LVI	Form B Item Data Obtained by Using Subtest Total Scores	352
LVII	Item Data for Pilot Test A	353
LVIII	Item Data for Pilot Test B	354
LIX	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals on Form A.	355
LX	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals on Form B.	356
LXI	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals for Grade X on Form A.	357
LXII	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals for Grade XII on Form A.	358

Table		Page
LXIII	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals for Grade X on Form B.	359
LXIV	Means, Standard Deviations, and Intercorrelations for Reading Test Scores and Poetry Subtest Scores and Totals for Grade XII on Form B.	360
LXV	Two-Way Analysis of Variance on Form A: Reading Level by Grade.	361
LXVI	Two-Way Analysis of Variance on Form B: Reading Level by Grade.	363
LXVII	Two-Way Analysis of Variance on Form A: Vocabulary Level by Grade	366
LXVIII	Two-Way Analysis of Variance on Form B: Vocabulary Level by Grade	368
LXIX	Two-Way Analysis of Variance on Form A: Comprehension Level by Grade.	371
LXX	Two-Way Analysis of Variance on Form B: Comprehension Level by Grade.	374
LXXI	Principal Components Solution and Varimax Rotation Transformation Matrix on Three Dimensions for Form A Pearson Item Intercorrelations.	376
LXXII	Principal Components Solution and Varimax Rotation Transformation Matrix on Three Dimensions for Form B Pearson Item Intercorrelations.	377

CHAPTER I

INTRODUCTION TO THE PROBLEM

BACKGROUND OF THE PROBLEM

From the time of Plato and Aristotle, much has been written about the nature and functions of poetry, but most of what is known about these matters has been provided by highly sensitive readers, many of them poets themselves. In order to teach poetry effectively teachers of English need a clear understanding not only of the nature of poetry itself but also of the interaction between a poem and its readers.

At present exact knowledge of what really happens when an ordinary reader encounters a poem is lacking. This kind of knowledge requires objective, systematic research of the kind pioneered by Richards and reported in Practical Criticism.¹ It is indeed important that English teachers be intimately conversant with the rich store of intuitive knowledge of poetry provided by writers and critics, and to have clearly in mind certain ideals of sensitive and mature poetry reading. However, without knowledge of the actual transaction between a poem and its reader, teachers can hardly

¹I. A. Richards, Practical Criticism (New York: Harcourt, Brace, 1929).

hope to be effective in leading their students toward a greater sensitivity to literary language and hence toward a deeper enjoyment of literature.

The need for objective research in reader response to literature is clearly stated by Ogden and Richards in

The Meaning of Meaning:

The person actually interpreting a sign is not well placed for observing what is happening. We should develop our theory from observation of other people, and only admit evidence drawn from introspection when we know how to appraise it. . . . Those who allow beyond question that there are people like themselves also interpreting signs and open to study should not find it difficult to admit that their observation of the behaviour of others may provide at least a framework within which their own introspection, that special and deceptive case, may be fitted.¹

At present among leading authorities in the teaching of high-school English there is widespread agreement that although one of the most complex problems in the teaching of English concerns the ways in which students comprehend literature while they read,² the classroom teacher continues to suffer from a dearth of systematic knowledge about this process. Gunderson states that the teacher of English gets little direct help from research in learning how to teach his students to become skilled readers of literature,³ and

¹C. K. Ogden and I. A. Richards, The Meaning of Meaning (New York: Harcourt, Brace, 1923), p. 26.

²Doris Gunderson, in Preface to J. C. Squire, The Responses of Adolescents While Reading Four Short Stories (Champaign, Ill.: N.C.T.E., 1964), p. v.

³Ibid.

in the same vein, Squire has recently remarked that "empirical research has not contributed notably to our understanding of the processes and procedures involved in literary education."¹

Closely allied to this problem of a lack of detailed knowledge about what Rosenblatt terms "the transaction" between a reader and a literary work² and largely contributory to it is the problem of a shortage of valid and reliable instrumentation with which to investigate reader reaction to literature. As Squire points out, "Until better ways of assessing response are available, we are not likely to stimulate many valuable and useful studies of the teaching and learning of literature."³ Gunderson, equally concerned with this problem of inadequate instrumentation, says, "When . . . we attempt to measure comprehension, we seem forced to simplify the content of material to be read so that we probably fail to get a measure of the individual's capacity to comprehend fully all that the author envisioned."⁴ And an even stronger statement of this view is made by Broening:

There is urgent need for sharper instruments for measuring the totality that is literary appreciation. New objective tests are required to diagnose the cause

¹Squire, The Responses of Adolescents, p. 56.

²Louise M. Rosenblatt, "The Acid Test for Literature Teaching," English Journal, XLV (February, 1956), 66.

³Squire, op. cit., p. 1.

⁴Gunderson, loc. cit.

or causes of a student's failure to appreciate literature appropriate to his level of intellectual and emotional maturity. Objectively determined criteria should be developed to evaluate what is read, how it is read, and what lasting impressions (provocative ideas, insights into motivation, enjoyment of verbal imagery, and appreciation of the power of language) remain with the reader.¹

The above comments refer to the twin problems of inadequate knowledge of student response and inadequate instrumentation for obtaining such knowledge as they relate generally to the teaching of literature. There is strong evidence that these problems become still more acute in respect to the teaching of poetry because of the relative complexity of poetic response. Richards, one of the first scholars to attack these difficulties in a systematic way, writes in Practical Criticism that a central purpose of his investigation was to advance the field of science, with its precision of measurement, into the field of poetry. Hyman, in discussing Richards' contributions to the criticism and teaching of literature, observes that

. . . the great contribution of Practical Criticism, and of Richards' work in general . . . is chiefly one of method and methodology. . . . By creating some approximation of experimental conditions, he achieves for the first time an objective account of how poetry is actually read, and focuses attention on its least adequately handled area, . . . the poem-reader relationship.²

¹Angela Broening, "The Development of Taste in Literature in the Senior High School," in N. B. Smith (ed.), The Development of Taste in Literature (Champaign, Ill.: N.C.T.E., 1963), p. 48.

²S. E. Hyman, The Armed Vision (New York: Vintage, 1955), p. 298.

Abbott and Trabue as long ago as 1923 were similarly concerned with the need for measures of response to poetry that would go beyond the mere acquisition of "a body of information, including information regarding conventional critical opinion. . . ." ¹ They saw the need for instruments by means of which "light would be thrown on the degree of success attained in developing the power to appreciate . . .," and by means of which "something might be learned of the general psychological response to the stimuli of the art in question, and so, of the fundamental principles of the art itself considered as a social activity designed to evoke popular response." ²

Loban, writing in 1947, is concerned about the inadequacy of instrumentation for measuring "growth in the study of literature," ascribing this inadequacy chiefly to the inherent complexity of literature itself. He quotes Smith's statement that

The development of educational measurement depends as much upon the nature of the phenomena with which it is concerned as upon the capacity of men to devise experimental procedures. For, as Aristotle observed long ago, we should not expect a greater degree of exactitude than

¹Allen Abbott and M. R. Trabue, A Measure of Ability to Judge Poetry (New York: Teachers College, Columbia, University, 1925), p. 3.

²Ibid.

the subject-matter will permit. We know the kind of subject-matter which measurement demands, but we do not know, nor can we ascertain, apart from experimental work, whether a given subject-matter has the systematic character required by measurement.¹

Because poetic language tends to be the most complex and intensified form of literary language, Loban's comment, while intended for literature as a whole, applies with double force to poetry.

Ironically, it is the very complexity of poetic language which has at the same time created for the English teacher a pressing need for more exact knowledge of the nature of student response to poetry and discouraged the kind of experimental work which might alleviate this need. Probably for the same reason, recent response studies, such as Squire's and Wilson's,² deal with fiction, not poetry.

The central assumption underlying this study is that approaches and techniques for the teaching of poetry must be determined not only by the nature of poetry and the reasons for teaching it, but also by the way the student actually responds to it. It is not enough for the teacher to know how he wishes the student to respond; at any given stage in the student's development, he must know how the student in fact does respond. Only with this knowledge can he hope to

¹Walter Loban, "Evaluating Growth in the Study of Literature," English Journal, XXXIV (November, 1947), 283.

²James R. Wilson, Responses of College Freshmen to Three Novels (Champaign, Ill.: N.C.T.E., 1966).

help students move toward a richer and more refined experience.

STATEMENT OF THE PROBLEM

The object of this study was to investigate the response of high school students to poetic language and to examine the usefulness of a procedure developed for measuring this response. The study sought to determine the kinds of information which could be supplied by an experimental poetry test constructed on the principle of matching individual words and expressions to the context of whole poems and supplemented by follow-up interviews based on responses to test items. The study attempted to identify and describe various components of student response to poetic language and to investigate relationships among these components. In addition, the study examined the relationship between response to various components of poetic language, as measured by the experimental poetry test, and response to "discursive" language as measured by a standardized reading test. Finally, the study attempted to determine relationships between grade level, sex, and reading level (as measured by a standardized reading test) and response to poetic language.

DEFINITION OF TERMS

Before stating the problem more fully by setting forth specific questions and hypotheses, it is necessary for

the sake of clarity to define the more important terms used in the study.

1. Poetry Test refers to the multiple-choice test constructed by the writer for this study.

2. Test Form refers to the two parallel forms of the Poetry Test, Form A and Form B, consisting of sixteen items each.

3. Reading Test refers to the Nelson-Denny Standardized Reading Test, Form A.

4. Follow-up Interview refers to the ten interviews which were based on responses to the Poetry Test.

5. Categories and Subtests refer to the four major divisions consisting of four items each, which comprise the Poetry Test. These categories are:

1. Thought
2. Suggestion
3. Sound
4. Rhythm

6. Skill in Response to Poetic Language refers to what is measured by the Poetry Test. (This term was used rather than such terms as "appreciation", "interpretive ability", "taste", and "sensitivity to language" because it is less ambiguous denotatively and more neutral connotatively than these other terms.)

7. Poetic Language, or Imaginative Language in its most intensified form, refers to the operation of words in a poem to produce a distinctively poetic meaning. It does not refer to a specialized poetic diction but is characterized by its suggestive quality and its emotive effects.

8. Discursive Language, used in opposition to Imaginative and Poetic Language, refers to verbal expression which is primarily referential or denotative, and is more concerned with the communication of ideas than of feelings.

9. Imaginative Language refers to language intended to arouse the imagination and emotions of the reader or listener. It attempts to cause him to recreate the original aesthetic experience of the writer, and is not primarily concerned with conveying fixed meanings.

10. Response refers to the active recreation of meaning arising from the interaction between a reader and a literary work.

11. Thought, or Denotative Meaning, refers to what is measured by Subtest One. It is substantially the same as what Richards calls the "thought" or "sense"¹ of a poem, and could be described as the "paraphrasable meaning" or "paraphrasable content" of a poem.

12. Suggestion refers to what is measured by Subtest Two. It is the meaning produced by the distinctive operations of the formal elements of poetic language, such as metaphor and imagery. For the purpose of this study, it does not include the effects created by the sound values and rhythms of words and groups of words.

13. Sound denotes what is measured by Subtest Three.

¹I. A. Richards, Practical Criticism, pp. 175, 200, 306.

For the purposes of this study it refers to the operation of segmental phonemes only and does not refer to rhythms and stress patterns. Sound, as virtually all authorities point out, rarely produces meaning independently: it typically acts in conjunction with the connotative or associational values of words. Hence Subtest Three consists of items in which sound was considered to make a significant contribution to the total meaning rather than items purporting to measure sound independently.

14. Rhythm refers to what is measured by Subtest Four. Rhythm was chosen rather than metre because, although most of the rhythm items involve metre, some do not. In certain of the rhythm items, for example, some of the alternatives are not metrical, while in others the metre has merely been changed so as to accord to a greater or lesser degree with the prevailing metrical pattern of the poem.

15. The names denoting certain formal elements of poetic language, such as imagery, have been used by critics and educators with a certain amount of ambiguity. For the purposes of this study, therefore:

- (a) Metaphor refers to all poetic comparisons, whether directly stated, as in the case of similes and metonymy, or implied, as in the case of symbolism.
- (b) Connotation refers to the use of words for their suggestive, evocative, or associative powers.
- (c) Imagery refers to the use of a word for its

powers of sensuous suggestion, whether visual, auditory, olfactory, or tactile.

(d) Ambiguity refers to the deliberate exploitation of the multiple denotative and connotative meanings of words and expressions.

16. Formal Elements, or Poetic Elements, is used in this study to refer to the various technical devices, used most consistently and characteristically in poetry, which distinguish imaginative from discursive language. These include such devices as imagery, metre, irony, ambiguity, and metaphor.

17. Dimension refers in this study to the components of skill in reading poetic language which are measured by the four subtests.

18. Rationalization refers in this study to the conscious awareness of the operation of the various formal elements of poetic language. The level of this awareness can vary: it might be low during the actual reading of a poem, and high when the poem is considered in retrospect.

19. Degree of Rationalization refers in this study to the extent to which a given act of selection can be recalled, brought to a conscious level, and coherently explained.

20. Quality of Rationalization refers in this study (a) to the degree of similarity between reasons actually underlying the original selection and reasons given later and (b) to the extent to which these original reasons

approximate those which might be expected from a highly qualified reader of poetry.

21. Correct is used in Chapter V to refer to the keyed response, which is the alternative taken from the original form of the poem.

QUESTIONS AND HYPOTHESES

The problem which this study seeks to investigate may be stated more specifically in terms of certain questions and hypotheses:

Questions Relating to the Performance of the Poetry Test

1. How reliable is the Poetry Test as a whole in terms of its purpose and in view of what it attempts to measure?
2. How similar are the two test forms? Are they statistically equivalent? If not, can they be considered comparable?
3. Is the difficulty level of the test appropriate to the group to which it is to be administered? Does it appear to be relatively more appropriate for the Grade XII group or the Grade X group?
4. What is the degree of correlation among scores on the four subtests? To what extent do they appear to measure discrete skills?
5. What patterns of correspondence, if any, exist among performance levels on the four subtests of the Poetry

Test? Is there, for example, a closer correspondence between performance on Subtests One and Two than there is between Subtests One and Three? Between performance Subtests Three and Four than between Subtests Two and Three? How does grade level affect these correspondence patterns?

6. How does performance on the four Poetry Test subtests correspond with performance on the Reading Test? Is this correspondence greater for some Poetry Subtests than for others? How does grade level affect this correspondence?

7. Can the testing technique used in the Poetry Test provide reasonably detailed information about the response to various elements of poetic language? Is it capable of measuring response to an adequately wide range of elements and in measuring this response at different levels of sophistication?

Hypotheses Relating to the Performance by Classes of Individuals

1. There is no significant difference between the mean performance of Grade X students and the mean performance of Grade XII students on Form A and Form B of the Poetry Test.

2. There is no significant difference between the mean performance of boys and the mean performance of girls for both grade levels and both test forms.

3. There is no significant difference between the mean performance on Form A and Form B of the Poetry Test

taken separately of students in the top third, middle third, and bottom third according to Reading Test scores at the Grade X level and the Grade XII level.

Questions Relating to Factors in
Response to Poetic Language

1. Does the pattern of factor loadings on the various Poetry Test items indicate the presence of a number of interpretable factors in the response to poetic language?
2. To what extent can these factors be identified and described?
3. How do these factors relate to other findings produced by this study, especially the interview results?

Questions Relating to Interview Results

1. To what extent are students generally able to account for selections made while responding to the Poetry Test?
2. Are students who make high scores on the Poetry Test better able to account for their choices than students who make low scores?
3. Do different students appear to use different strategies of thinking in the way they attack the various items?
4. Are students better able to account for selections within items from some of the subtests than for selections within items from other subtests?
5. How stable are the test items, i.e., how

accurately are students able to recall the selections made?

6. Are item stability, and test stability as a whole, related to proficiency in answering the Poetry Test items?

7. What is the role of context in the making of the selections? Does this role differ among the different kinds of items, i.e., among items from the different subtests?

8. To what extent are interview results related to the results produced by factor analysis? To what extent do the interviews lend support to factors which have been identified and described?

DESIGN OF THE STUDY

In order to investigate the above questions and hypotheses an investigative procedure consisting primarily of an experimental poetry test was developed. This test classified poetic language under four categories designated as "Thought", "Suggestion", "Sound", and "Rhythm". Two forms, each comprising 16 items, four in each category or subtest, were developed. The testing procedure involved presenting the student with a whole short poem, or reasonably integral section of a larger poem, from which a word or phrase had been deleted, and requiring him to select the original word or phrase from four alternatives using "clues" from the context.

Supplementary to the poetry test itself, a follow-up interview programme was developed. Essentially, the

interviews were structured in order to require students to explain their selections within the various test items, but provision was also made for a certain measure of free response.

The test items, after content validation by four judges, were used in a number of pilot studies so as to obtain optimal difficulty levels and a high degree of discriminating power for each item. The interview schedule was also tested by means of a pilot study.

The resulting two forms of the poetry test were then administered in conjunction with the Nelson-Denny Reading Test, Form A, to eight classes of high school students in North Vancouver, B. C., four from Grade X and four from Grade XII. These classes, consisting of 191 students taking the academic programme, were selected at random from a population of approximately 80 classes and 2400 students. Within seven days following the administration of the Poetry Test, interviews were conducted with a representative subsample of ten students.

In order to examine the reliability of the poetry test and to determine its usefulness in providing information about various aspects of students' response to poetic language, test results were subjected to item analysis. Intercorrelations were obtained for the Poetry Test totals and subtotals and the Reading Test totals and subtotals. To compare relationships between grade level, sex, and reading level and Poetry Test performance two-way analyses of

variance were carried out. A principal components factor analysis was performed on Poetry Test results in an effort to examine the underlying factor structure of poetic language.

In addition, the recorded interviews were analyzed in order to obtain information about students' thought processes while reading poetic language which would supplement the results of the statistical analyses of test data.

DELIMITATIONS OF THE STUDY

1. In view of the improbability of obtaining a valid measure of affective response to poetic language through a pencil-and-paper test, the Poetry Test purports to measure directly only cognitive aspects of response. "Appreciation" is examined only indirectly.

2. The Poetry Test emphasizes skill in responding to poetic language rather than skill in deriving a global meaning from a whole poem.

3. The design of the experimental test required the reduction of a large number of poetic elements into a relatively small number of categories (four). Therefore, the study does not attempt to measure equally all important aspects of response to poetic language.

4. The sample of subjects was taken from a population of Grade X and Grade XII students. The findings, therefore, may have only limited application to other educational levels.

LIMITATIONS OF THE STUDY

1. The small number of items (four) within each category, or subtest, which was made necessary by the test design, limits the generalizability of conclusions arising from responses to any one of the subtests.

2. Because the Poetry Test was not a one-factor test, because a test-retest procedure was not feasible, and because it was not possible to administer both test forms to the same subjects, only a partial reliability can be demonstrated for the present test.

3. As the interviews were conducted by the experimenter, responses relating to individual items were possibly subject to distortion due to the interviewer's expectations.

4. The interview results were dependent on such factors as the student's ability to recall thought processes and the student's verbal fluency. Because of problems in establishing credibility and in detecting spurious rationalization, the interview results need to be interpreted with caution.

SIGNIFICANCE OF THE STUDY

As was reported earlier there is general agreement among authorities in the teaching of English that the more effective teaching of poetry entails showing students how to derive meaning from "poetic" as opposed to "discursive" language. The teaching of poetry can be regarded as the

teaching of a special kind of reading skill requiring on the part of the teacher accurate information about the way students respond to the language of poetry. This kind of information is lacking largely because of a serious lack of the kind of instrumentation which reveals the underlying factors in response to poetic language and which provides objective knowledge about a student's thought processes as he reads a poem.

The present study accordingly attempts to develop and assess the value of a specific procedure for investigating student response to poetic language. At the same time the study attempts by means of this procedure to throw light on some of the factors underlying this response. It is hoped not so much that definitive conclusions will be reached but that the investigative procedure used will stimulate other and perhaps more sharply focused research into student response to poetic language.

PLAN OF THE STUDY

Chapter II presents a review of related literature. First an examination of statements by critics and educators regarding the nature of poetic language is presented. This is followed by a review of statements regarding the testing of literature in general, the measurement of student response to literature, and the measurement of student response to poetry in particular.

Chapter III sets out the research design for the

study. In it are described the development of the experimental poetry test and the interviewing procedure, the sampling procedure and collection of data, and the statistical treatment of the data.

Chapter IV presents the results and interpretation of the statistical analysis and the discussion of these results.

Chapter V presents the findings derived from the interviews, together with a discussion of these findings as they relate to the results reported in chapter IV.

Chapter VI contains a summary of the main findings of the study, generalizations resulting from the study, suggestions for further investigation, and certain implications for the teaching and testing of poetry.

CHAPTER II

SURVEY OF RELATED LITERATURE

INTRODUCTION

Following upon the basic assumption that the primary emphasis in the teaching of poetry must be on the student's ability to respond adequately to the language of poetry, the present chapter is mainly devoted to examining the nature of poetic language.

Because of a lack of empirical investigation there is a current shortage of systematic, objective information about a student's thought processes as he reads a poem. Thus knowledge of the difficulties likely to be encountered by a high school student in reading a poem must for the greater part be inferred from what has been written by artists and scholars about the characteristics of poetry and poetic language. This review accordingly deals mainly with the writings of poets, critics, literary theorists, and authorities in the teaching of literature.

Because the construction of an objective test forms a major part of the study reference is made to the testing of literature, with special emphasis on the testing of poetry. Also examined are studies of student response to literature in general and to poetry in particular.

THE NATURE OF POETIC LANGUAGE

The main purpose of this section is to review certain commonly agreed-upon features of poetic language which distinguish it from discursive language and which are most likely to account for a student's difficulties in reading a poem. The term "poetic language", as used here, is to a large extent synonymous with "imaginative language" and the essential distinction is between "poetic" and "discursive" language.

Generally speaking the reader's difficulties stem from two sources: first, the inherent complexity of poetic language, with its many dimensions of meaning, and second, the difference between the ways in which poetic and discursive language embody and transmit meaning.

Both this inherent complexity and this difference are due mainly to what linguists term the "hypersemantised" aspect of poetic language, i.e., its capacity to convey meanings over and above those connected with normal syntactic and semantic rules. These extraordinary meanings result first, from the poet's concern with the total meaning of a word. Unlike the writer of a scientific report the poet is concerned with a word's sensuous qualities, its evocative power, and its sounds.

Second, poetic meanings result from the organic unity of a poem, by which is meant not only that structural elements act interdependently in a poem to produce meaning but also

that in a very real sense these structural elements are the meaning. Even in strictly denotative language words tend to have a number of possible meanings, and as a rule the greater the possibilities for ambiguity, the greater the importance of context in guiding the reader or listener in choosing among these possible meanings. In poetry, with its many types of meaning, this process of selection becomes still more demanding and complex. Hence in poetic language the role of context is of maximum importance. In poetry the bond between one word and all the other words of a poem is stronger and consists of more strands than the corresponding bond in discursive language. This intimate and complex bonding is essentially what is meant by "the organic unity" of a poem.

Accordingly, this section reviews critical statements pertaining to characteristics of poetic, or "imaginative", language which distinguish it from discursive language. Examined first are general characteristics, such as the "organic unity" of poetic language and the multiplicity of meanings found in poetry. This is followed by a more detailed examination of poetic language as it relates to the main structural elements of poetry: connotation, metaphor, imagery, and sound. Although there is some unavoidable overlap between the first, more general, and the second, more detailed, review, each treats poetic language from a different perspective.

General Characteristics of Poetic Language

An examination of English literary criticism for views on the general nature of poetic language reveals two focal figures, Pope and Coleridge. Pope, reflecting the Horatian, or neo-classical tradition, and less directly the classical view that art is an imitation of persisting, objective forms in nature, sees poetic style as related to but still separable from the matter of the poem. For him poetic language is "the dress of thought",¹ with the sound "echoing the sense".² For him language as it operates in a poem is essentially an extension and refinement of discursive language. The poet by means of his special skill with words merely expresses a meaning or reality which is external to the poem.

Coleridge on the other hand sees the matter of a poem as living only in its words and as non-existent apart from its language. In Coleridge's view the function of art is to imitate nature not by merely imitating the concrete particulars of the natural world or even through selection among these particulars to reveal universal forms. What art seeks to imitate rather is the process in nature through which concrete particulars and universal forms fulfil each other and exist only through one another. The creation of a work of art is a re-enactment of this process,

¹Alexander Pope, "An Essay on Criticism," 1.312.

²Pope, op. cit., 1.365.

an incarnation of reality. Hence in a poem the reality which concerns poet, critic, and reader alike exists in the interaction between the concrete particular and the overall form--between the whole and its parts. Thus the poem, according to Coleridge's view, is conceived of as an organic unity where the whole is far more complex than the sum of its parts. Through "multeity", or "a fusion to force many into one",¹ a poem duplicates the active principle of nature.

This well-known Coleridgean view has been presented at some length because generally speaking it underlies most of the currently prevailing ideas on the operation of language in a poem. The influence of Coleridge can easily be traced through Richards and the "New Critics" to most of the current works on the nature and the teaching of poetry.

The following description of general characteristics of poetic language therefore is closely related to the Coleridgean view. Each of the characteristics is in fact merely a facet of the overall idea of "organic unity". Accordingly these characteristics are closely related.

Organic unity. According to the Coleridgean view, the meaning of a poem is produced by a harmonious involvement of all of its parts. It is in the process of

¹S. T. Coleridge, On the Principles of General Criticism: Essay Third, in W. S. Bate (ed.), Criticism: The Major Texts (New York: Harcourt, Brace, 1952), p. 369.

interaction among all the various elements that the meaning resides and just as the poet re-creates, or "abridges",¹ reality by imitating the active process of nature in which the universal and the particular fulfil each other the reader, in turn, is actively involved in the re-creation of reality during the act of reading. Only through this re-creation of reality can the reader derive meaning from the poem.

Thus, in poetic, as opposed to discursive language, the meaning of a word or expression or the effect of any one structural element is far more closely dependent on context and the meaning of one part is more intimately connected with the meanings of all the other parts, both separately and in combination.

Most modern critics, perhaps because of the influence of I. A. Richards, tend to echo Coleridge in this matter of organic unity. It is commonly recognized that the more successful a work of art the more its substance becomes inseparable from its form. Ciardi, in How Does a Poem Mean? notes:

For What Does a Poem Mean? is too often a self-destroying approach to poetry. A more useful way of asking the question is How Does a Poem Mean? Why does it build itself into a form out of images, ideas, rhythms. How do these elements become the meaning? How are they inseparable from the meaning? As Yeats wrote, "O body swayed to music, O quickening glance/How shall I tell the dancer from the dance?" What a poem is, is inseparable from its own performance of itself.²

¹Bate, op. cit., p. 369.

²John Ciardi, How Does a Poem Mean (Boston: Houghton-Mifflin, 1959), p. 667.

In much the same vein, Diltz speaks of a poem as a whole made up of many parts with "each word . . . entangled in the meaning, suggestion, feeling, imagery, and rhythm of the whole,"¹ so that the language takes on a significance far beyond its dictionary meaning.

Also subscribing to this Coleridgean view of the whole being a harmonious involvement of all of the parts are such writers as Phenix,² Gurrey,³ Guth,⁴ Richards,⁵ Scott,⁶ and Murphy.⁷ R. P. Warren's statement that "Poetry does not inhere in any particular element but depends upon the set of relationships, the structure, which we call the poem"⁸ is perhaps the most succinct comment on this whole matter, not only of the unity of form and

¹B. C. Diltz, The Poetic Pilgrimage (Toronto: Clarke Irwin, 1942), p. 667.

²Philip H. Phenix, Realms of Meaning (New York: McGraw-Hill, 1964), p. 178.

³P. Gurrey, The Appreciation of Poetry (London: Oxford University Press, 1935), pp. 56-63.

⁴Hans Guth, English Today and Tomorrow (Englewood Cliffs: Prentice-Hall, 1964), pp. 282-84.

⁵I. A. Richards, Practical Criticism (New York: Harcourt, Brace, 1929), pp. 173, 328.

⁶Wilbur Scott, Five Approaches to Literary Criticism (New York: MacMillan, 1963), *passim*.

⁷Geraldine Murphy, The Study of Literature in the High School (Toronto: Blaisdill, 1968), pp. 16-38.

⁸R. P. Warren, Selected Essays (New York: Random House, 1958), p. 41.

content but of the interdependence of the formal elements themselves.

Because of the close interdependence of structural elements and of individual words and phrases which "organic unity" implies, the meaning of a poem can emerge only from a rapid and constant shifting of the reader's attention among its various parts. Coleridge speaks of the need for ". . . the perpetual and distinct attention to each part,"¹ and notes that poetic, as opposed to discursive language, "is discriminated by proposing to itself such delight from the whole as is compatible with a distinct gratification from each component part."²

The difficulty for the student in this matter is that he is habituated in his "normal" use of language to consider "content", or "sense", or "meaning", as something separate from the language which conveys it, so that when he approaches a poem he tends to bring a faulty set of expectations as to the kind of meaning he should be looking for. If, with Pope, he sees poetic expression as merely "the dress of thought", i.e., as merely a refined version of discursive language, and attempts to extract meaning from words and expressions using only standard syntactic and lexical rules and neglecting the "distinct and

¹Bate, op. cit., p. 378.

²Ibid.

perpetual attention" to the relation between each part and its context required by poetic language, he is likely to be baffled and disconcerted. What he fails to realize of course is that a poem, as an organic unity, demands his active participation and cannot be treated as an artfully constructed or cleverly embroidered piece of discursive writing, expressing "what oft was thought, but ne'er so well expressed."¹ According to the Coleridgean view, which is essentially the one currently prevailing, the reader of a poem must be actively engaged in the creation of meaning, not through imposing his own abstractions on the poem, but by using his experience to allow him to perceive universal forms emerging through its concrete symbols.

Fluidity of meaning. As has been noted above the notion of organic unity implies a close interrelationship between the words and structural elements of a poem, with the meaning of each part being far more dependent on its context than is the case in discursive language. An obvious corollary is the notion that words and expressions as they operate in poems tend to have fluid rather than fixed meanings.

In recent years linguists and psychologists have pointed out what many writers and critics have known all along, i.e., that word meanings are not fixed and static,

¹Pope, op. cit., 1.297.

to be learned once and for all and accumulated in the mind, but are forever modified by the context in which they are used and by the level of awareness of the individuals using the language.

Vygotsky defines the "sense" of a word as "the sum of all the psychological events aroused in our consciousness by a word" and describes it as "a dynamic, moving and complex whole."¹ According to Vygotsky, "meaning", unlike "sense", is "only one of the zones of that sense which is acquired by the word in the context of speech, namely, the most stable, unified, and precise zone."² Vygotsky's view that word meanings develop, both in the consciousness of individuals and in the consciousness of the race, is of fundamental significance in understanding the nature of poetic language. His theory makes it possible to conceive of poetic language not, as many linguists have done, as an aberration from normal language, but as a phenomenon which relates to two fundamental, opposed, and complementary human needs. These are the need for the development of individual consciousness and the need for social intercourse. Vygotsky sees in the course of the historical development of language "the structure itself and the psychological nature of meaning" changing in the direction

¹L. S. Vygotsky, Thought and Language (Cambridge, Mass.: M.I.T. Press, 1962), p. 530.

²Ibid.

of greater generalization and abstraction in the expression of ideas. At the same time he sees as developing in the child something he calls "inner speech" in which "the semantic aspect develops from the whole to the particular, from sentence to word."¹ It would seem that paralleling this process within the child and in opposition to the movement of language toward greater abstraction is the poetic use of language, where the direction is toward concreteness and particularity.

This fundamental opposition between language as transmitter of meaning (discursive language) and language as creator of meaning (poetic language) has been noted by other psychologists and by many linguists, semanticists, and literary critics. Barfield sees the main function of poetry as either the creation of new meanings or the recreation of lost meanings, "pointing up the resemblances between things."² Richards, who in his earlier writings distinguished between "emotive" and "referential" language, later came to see a more useful distinction as between "rigid" and "fluid" language:

There is an important use of words - very frequent, I suggest, in poetry - which does not freeze its meanings but leaves them fluid, which does not fix an assertorial clip upon them in the way that scientific

¹Vygotsky, op. cit., p. 1515.

²Owen Barfield, Poetic Diction (London: Faber and Faber, 1926), p. 102.

prose and factual discourse must. It leaves them free to move about and relate themselves in various respects to one another.¹

Ullman, while noting that there is usually a hard core of meaning which is relatively stable in a word, and which can only be modified by the context within certain limits, also recognizes this fluidity in linguistic meaning, and notes that "the influence of context increases with possibilities of ambiguity."² Because it is in poetic language that these possibilities of ambiguity are most evident, this view seems to be in complete accord with Richards' fluidity-rigidity continuum.

These statements all reflect an awareness of linguistic meaning as "fluid", or "dynamic". They also point to the fact that the degree of "fluidity" of a given word varies according to the kind of utterance of which it forms a part; and that words achieve their greatest degree of "fluidity" in poetic language. Further, these statements point to an awareness of two opposing forces in language development, both of which are necessary to human development. One of these, seen best in purely discursive or referential language, is toward fixed meanings, toward generalization and abstraction; the other, characterized by

¹I. A. Richards, Speculative Instruments (London: Routledge and Paul, 1955), p. 148.

²Stephen Ullman, Semantics (Oxford: Blackwell, 1962), p. 48.

poetic language, is toward synthesis, toward expressing as much meaning as possible in a single word or expression. The object of the first of these is to fix meanings so as to make efficient communication possible; the object of the second is to create new meaning through the medium of language.

Thus it is clear, as Buyssens¹ points out, that each of these two forces is necessary for the other's health and even survival. The artist is primarily interested in expression, but in time, if he is sufficiently skilled and his audience sufficiently responsive, he can achieve richer and more precise communication. Conversely a keen awareness of fixed rules of syntax and semantics, of the "hard core of meaning" to which Ullman refers, is as necessary to the artist creating new meanings by opposing or stretching them as it is to the purely discursive speaker whose sole object is to comply with them.

In Chomsky's phrase, literature "makes strange"² with language. The suggestion is that poetic meaning is generated through a tension between established language usage and the "new" mode of utterance. It would seem that the greater the sensitivity to the refinements of

¹Eric Buyssens, La Communication et L'Articulation Linguistique (Paris: Presses Universitaires de France, 1967), pp. 23-24.

²Noam Chomsky, Language and Mind (New York: Harcourt, Brace and World, 1968), p. 21.

established usage, the greater the possibilities for the production of poetic meaning.

The preceding section has examined the relationship between poetic and ordinary language in terms of the notions of "fluidity" and "rigidity". It has presented the notion of poetic language as lying at the opposite end of a linguistic continuum from purely discursive, or referential, language, and the notion that the words of poetry are fluid, as opposed to rigid, in their meanings.

Multiplicity of meanings in poetic language.

Closely related to the view that poetic language is characterized by fluidity of meaning is the notion of "multiple meaning". Perhaps the greatest source of difficulty for the uninitiated reader in approaching a poem is the characteristic tendency of poetry to use words and expressions to convey separate and often contradictory meanings simultaneously. Empson, in his influential Seven Types of Ambiguity,¹ sees imprecision of meaning not as a deficiency but as the chief virtue of poetry: he finds ambiguity the essential mark of poetic expression. Not only can a word or expression have several meanings in one dimension, such as the denotative dimension; it can also have one or more meanings in yet another dimension. Thus, the sound of a word might simultaneously reinforce its denotative meaning

¹William Empson, Seven Types of Ambiguity (New York: Meridian, 1955), p. 12.

and run counter to its connotative meaning.

Ciardi notes that while all language is to some degree characterized by the simultaneous operation of different meanings, poetry differs from ordinary language in deliberately and consistently exploiting ambiguity. He notes that the poet simply makes his selections with greater attention to the various forces within the word.¹

Here again the implication for the teacher of English literature is that the reading of poetry must be an extremely complex process. It is in the tendency of poetry to make use of multiple meanings that much of its power and unique value resides. If used with skill poetic ambiguity, as Empson shows, allows the poet to exploit the full power and subtlety of his language. In controlling the relationships and tensions between the various meanings of a word or group of words he can express what could not be expressed otherwise.

Perhaps the great value of multiple meanings is that they make possible the organic unity the poet is attempting to achieve by forcing the reader to become involved in the poetic process and to form his own synthesis of the total meaning of a poem. In effect the reader is compelled to re-create the poem. Early, stressing the importance of the transaction between poem and reader in the poetic process, speaks of the reader as struggling for the artist's

¹Ciardi, op. cit., p. 763.

meaning."¹ Richards, in one of the clearest statements of the effect of multiple meanings, points out that

There is a productive activity between the words as we read them. Following, exploring, becoming that activity is, I suggest the essential thing in reading the poem. Understanding it is not a preparation for reading the poem. It is itself the poem. And it is a constructive, hazardous, free creative process, a process of conception, through which a new being is growing in the mind. . . . We seem to create [the] poem.²

The more demanding the synthesis, therefore, the greater should be the reader's involvement with the poem. It is not merely, as Richards' earlier work suggests, a case of transmitting a ready-made experience. The meaning is rather the set of interactions between poet and poem and between poem and reader.

Another value of poetic ambiguity is that the poet, in manipulating the way the many kinds of meaning at his disposal reinforce and oppose one another, can achieve not only a high degree of subtlety and precision but a richness of meaning not possible with discursive language. At the same time the use of multiple meaning can produce a degree of vigour and intensity rarely found in discursive language. Equally important, this concentration of expression makes possible the quality which Aristotle calls perspicuousness.³

¹Margaret Early, "Stages of Growth in Literary Appreciation," English Journal, (October, 1960), 164.

²Allen Tate (ed.), The Language of Poetry (Princeton: Princeton U. Press, 1942), p. 76.

³Bate, op. cit., p. 24.

This quality, which is made possible only through the conciseness produced by multiple meanings, allows the poem to be seen as an organic whole.

It will be clear from the above statements that this tendency of words to mean several things in several ways at the same time is an essential characteristic of poetry and of the poetic process. Combined with the characteristic "fluidity" of poetic language discussed earlier, it produces the quality of suggestiveness so prized by the romantic poets and critics and so typical of modern poetry. As Coleridge notes, and as has been suggested in previous paragraphs, the power of poetry is not attained through detailed description but by forcing the reader's imagination to create the picture.

It will also be clear that the reader who approaches a poem expecting to find unambiguous, fixed meanings will experience considerable difficulty. The implication here for the teacher of poetry is that the main task is to make students aware of the way poets deliberately exploit ambiguity in their creation of poetic meaning.

Poetic language and poetic diction. At this point it seems necessary to deal with a possible difficulty regarding the use of the term "poetic language". At no place in this study is this term intended to denote "poetic diction", i.e., a specialized vocabulary peculiar to poetic utterance. Throughout the history of literary criticism,

there has been considerable controversy over whether the language of poetry differs from the language of ordinary or "discursive" speech in kind or merely in degree. This study, in accordance with the broad consensus of current critical opinion, takes the view that a word or expression is "poetic" not in itself but in the special way in which it is used to convey poetic meaning. Richards, supporting his basic contention that literature is essentially a matter of communicating experience by means of a certain aesthetic mode of expression, contends, as did Wordsworth and Coleridge, that the development of a separate, specialized poetic diction makes true poetic meaning impossible and that "poetic meaning" can exist only in terms of the way poetry is interpreted. As Schiller, in a recent review of Richards' contribution to criticism, points out:

Richards requires for his literary theory a theory of interpretation with which he can locate the broad contrast between poetic and non-poetic interpretation, not a theory of meaning. He needs a commonsensical contrast between our poetic and non-poetic activities. . . . The writer shares his medium with all who use the language. The words he uses are the same as those used in everyday discourse. He is, in fact, confined to the words and to the meanings of each that native speakers have agreed upon. There is no literary vocabulary; there are no "poetic" words. The writer simply makes ordinary words behave in extraordinary ways.¹

There is also a need to be clear about the relationship between "poetic language" and the language of literature

¹J. P. Schiller, I. A. Richards' Theory of Literature (New Haven: Yale University Press, 1969), p. 147.

as a whole. Here again the difference is one of degree rather than of kind. Perhaps it is most useful to conceive, as does Murphy, of a continuum in the use of language, with purely discursive or referential speech at one end, lyric poetry at the other, and other literary genres in an intermediate position. According to this view, which accords with the position taken by this study, poetry, and especially lyric poetry, merely utilizes the resources of language more fully and in a more concentrated fashion than do other types of literature.¹

Summary. The overall implication of the above comments on the general nature of poetic language for the teaching and hence testing of poetry is that the central task is to give the student a clear understanding of the nature of poetic language in terms of its purposes and to give him a clear understanding of the differences between the ways in which poetic and discursive language operate.

The chief implication of the preceding discussion for the present study is that teachers require accurate means of measuring the degree to which students achieve the "clear understanding" referred to in the previous paragraph. If the acceptance of the Coleridgean idea of "organic unity" is as widespread among poets, critics, and teachers as a survey of the relevant literature indicates,

¹Geraldine Murphy, op. cit., p. 63.

the testing of poetry ought to reflect this idea; thus testing must focus on the student's ability to recognize how the individual parts of a poem relate to their context. A valid test of poetry must require "distinct and perpetual attention" to the various parts in order to arrive at an understanding of the ways in which they relate to each other and to the whole poem. Unlike the many poetry tests which treat these parts in isolation and appear to make no distinction between poetic and discursive language, the present test attempts to deal with poetry on its own terms. That is, in focusing on the process by which the parts interact with the whole, it proceeds on the assumption of "organic unity".

Specific Structural Elements of Poetry

Over the centuries poets and critics have supplied an abundance of material on the nature of poetry and the way its various structural elements such as rhythm and metaphor contribute toward the total meaning of a poem. It is neither necessary nor possible for a review such as this to deal with all these elements or even with any of them with any degree of completeness. What is attempted in this section is rather to review statements by authorities on poetry concerning four of these structural elements (connotation, imagery, metaphor, and sound) which seem most relevant to the central problem of this study: that of learning how to respond adequately to poetic language.

Connotation. The words of a language, aside from their denotative, or dictionary meanings, carry with them an accumulated suggestive, or connotative meaning.¹ Vygotsky, speaking of language in general, points out that, "The sense of a word is the sum of all the psychological events aroused in our consciousness by the word,"² and Ogden and Richards, also concerned with enlarging the concept of linguistic meaning, note much the same thing:

Words or arrangements of words evoke attitudes both directly as sounds, and less directly in several ways through what are called loosely "associations" More important than the sound qualities of words are the immediate emotional accompaniments due to past experience of them in their typical connections. To get these, there is no need for the connections themselves to be recalled.³

Thus words, in varying degrees, take on feelings, whether used in poetry or in everyday discourse. The difference between the way connotative meanings operate in normal speech and in poetry is, as Ciardi points out,

simply that the poet makes his selections with more acute attention to the various forces within the word. For the person who is a language-sensitive, every word has its own personality . . .; every word releases its aura of feeling.⁴

The tendency of words to gather emotional associations from everyday experience is perhaps the chief reason

¹Ciardi, op. cit., p. 762.

²Vygotsky, op. cit., p. 530.

³Ogden and Richards, The Meaning of Meaning, p. 373.

⁴Ciardi, op. cit., p. 763.

for the power of poetry. T. S. Eliot, noting the symbiotic nature of the relationship between poetry and everyday language, points out that

The poetry of a people takes its life from the people's speech and in turn gives life to it, and represents its highest point of consciousness, its greatest power and its most delicate sensibility.¹

But connotation is also the source of much of our difficulty with poetry: because of its centrality in poetic language, non-native speakers find it almost impossible to appreciate the poetry of a language other than their own. As Richards says, it makes poetry untranslatable. It also determines the way the poet must operate, and raises a problem. To exploit the full evocative power of language, the poet must lean heavily on what Conrad calls "the commonplace surface of words, of the old, old words, worn thin, defaced by ages of careless usage."² He must, however, use these "old, old words" in creating new meanings. If he yields too much to the temptation of coining new words, or of creating a specialized diction, he fails to tap the deepest feelings of his readers. So important to the total poetic effect is connotative meaning that poets give great care to establishing the right relationships between the

¹T. S. Eliot, *The Use of Poetry and the Use of Criticism* (London: Faber and Faber, 1933), p. 15.

²Joseph Conrad, *The Nigger of the Narcissis: Preface* (New York: Vintage, 1965), p. ix.

connotation of the words in a given poem.¹

Connotation is closely related to the other dimensions of poetic meaning. Imagery and metaphor, in the sense used in this study, can be considered as merely specialized forms of connotation. The dimension of word sound, however, is the one most closely connected to connotation. According to Ogden and Richards,

The means by which words may evoke feelings and attitudes are many. . . . As sounds and again as movements of articulation, as well as through many subtle networks of associations, the contexts of their occurrences in the past, they can play very directly upon the organized impulses of the affective volitional systems.²

Ullman, a semanticist, also notes the closeness of the relationship between the sounds and associations of words,³ and Ciardi notes that most people respond instinctively to the musical qualities of words.⁴ Ciardi's view is that connotation presents less of a problem in teaching than imagery and metaphor, because its use in poetry coincides more closely than theirs with its use in normal speech.

Imagery. Literary critics have often used "imagery" to include all uses of figurative and metaphoric language. In this study, "imagery" denotes the representation through language of sense experience. Though it most often

¹Ciardi, loc. cit.

²Ogden and Richards, op. cit., p. 377.

³Ullman, op. cit., pp. 83-89.

⁴Ciardi, op. cit., p. 766.

suggests mental pictures, imagery is not confined to the visual: it may represent a sound, a smell, or a touch.¹

Imagery is particularly important to the poet because of its power to evoke experience and to arouse emotions through producing mental reproductions of sensation.² Most authorities on poetry are in substantial agreement regarding the importance of imagery.^{3,4,5,6} Spender, speaking from personal experience, insists that

The poet, above all else, is a person who never forgets certain sense impressions which he has experienced and which he can re-live again and again as though with all their original freshness.⁷

And Conrad, though referring specifically to fiction, voices his belief in the central role of imagery in literature:

My task which I am trying to achieve is, by the power of the written word to make you hear, to make you feel - it is, before all, to make you see. That - and no more, and it is everything.⁸

Clearly imagery is closely related to the other dimensions

¹Laurence Perrine, Sound and Sense (New York: Harcourt, Brace, 1956), p. 45.

²Perrine, op. cit., p. 48.

³Ciardi, op. cit., p. 866.

⁴Cleanth Brooks and R. P. Warren, Understanding Poetry (New York: Holt, Rinehart, 1938), p. xiii.

⁵Guth, op. cit., p. 286.

⁶Murphy, op. cit., p. 32.

⁷Stephen Spender, The Making of a Poem (London: Hamish Hamilton, 1955), p. 55.

⁸Conrad, op. cit., p. ix.

of poetic meaning examined in this study and can perhaps be considered as a special kind of connotative language. Many critics even go so far as to lump all poetic comparisons, such as metaphor, under the heading of imagery, but for the purposes of this study "imagery", in line with the above statements, will denote concrete language making its primary appeal through the senses.

The ordinary reader's problem in responding adequately to imagery is that he has been conditioned to respond primarily to abstract, discursive language. Ciardi refers to a word's "picture-making power" as an indispensable function in determining the feeling of a word. He says, "Words are most richly conceived as individuals, each with its own history, each with its own locked-in picture."¹ However, as Barfield points out, this "picture-making" function of a word is countered by an opposite, often stronger, force, "by which . . . single meanings tend to split up into separate and often isolated concepts."² In this continuing war between unity of meaning, one of whose chief defenders and exemplars is poetry, and fragmentation and abstraction of meaning, clearly evident in the language of science, the latter appears at the present time to be winning. The uninitiated reader of poetry, caught in the middle, cannot be blamed for "groping about" as Eliot puts

¹Ciardi, op. cit., p. 766.

²Barfield, op. cit., p. 87.

it "for what is absent . . . for a kind of 'meaning' which is not there, and is not meant to be there."¹ He has been taught to look for a meaning which is now evidently missing.

Metaphor. For the purposes of this study "metaphor" is taken to be synonymous with "figurative language" and includes the use of simile, personification, and metonymy. To a great extent it also includes the use of symbol, both metaphor and symbol being subsumed under what Ciardi terms "the metaphoric sense". Emphasizing the centrality of metaphor in poetic language, he says,

Nothing is more characteristic of poetry than this metaphoric sense. . . . Literal-minded readers . . . are forever in difficulty with the metaphoric sense of poetic language. They want to know "what a poem is about," and they resent the essential duplicity whereby the poem is never entirely what it seems to be about.²

This "essential duplicity" of metaphoric expression is also noted by Perrine, who talks of "saying one thing and meaning another" as a means of adding extra dimension to language,³ and by Walsh, who says,

Poetic indirection is not an optional adornment. It is a method of putting into words, with exact fidelity, ideas and emotions and moods that might otherwise elude the capacities of language.⁴

¹Eliot, op. cit., p. 150.

²Ciardi, op. cit., p. 866.

³Perrine, op. cit., p. 54.

⁴Chad Walsh, Doors into Poetry (Englewood Cliffs: Prentice-Hall, 1962), p. 61.

It is now necessary to look more closely at the way in which metaphor achieves its remarkable effectiveness. It was noted earlier that there appears to be a natural tendency in the development of language for single meanings to become fragmented into isolated and separate concepts or abstractions and that perhaps one of the chief functions of poetry is to oppose this process, creating or restoring unified, rich, complex meanings. In a previous section dealing with the multiplicity of meanings in poetic language it was noted that deliberate ambiguity or plurisignation is one effective means of expressing large and complex meanings. Perhaps even more effective in this struggle is the use of metaphor. Richards sees the distinguishing mark of the poet in his ability to make original and apt connections between different elements of his experience,¹ and R. P. Warren states that

. . . other things being equal, the greatness of a poet depends on the extent of the area of experience which he can master poetically.²

One of the most effective devices in the making of these connections and in the "mastering" of a wide area of experience is metaphor. But metaphor does not merely act as a device for gathering and combining a large number of fixed meanings. It allows the poet, tied as he is to the unyielding words of his language, to fulfil his central

¹Richards, op. cit., p. 127.

²Warren, op. cit.,

function of creating fresh meanings through combining words and ideas in new ways. As Barfield notes,

the principal means by which creation of meaning is achieved is . . . metaphor. But it must be remembered that any specifically new use of a word or phrase is really a metaphor, since it attempts to arouse cognition of the unknown by suggestion from the known.¹

Coleridge, who has provided some of the most perceptive and influential insights into the relationship between poetry and the human imagination, sees imagination not merely as a picture-making faculty, but as the agent which enables the human mind to conceive simultaneously of universal forms and the concrete, particular world. As Bate says in discussing Coleridge, "Imagination, expressing itself through poetry, 'fuses' the insight of reason with the impressions of the senses."² Clearly metaphor is one of the chief means for effecting this fusion.

But in arousing cognition of the unknown by suggestion from the known metaphor, far from simplifying meanings, far from handing the reader a ready-made relationship, forces the reader to participate in the creation--or re-creation--of meaning. Like ambiguity, metaphor not only allows for multiple meaning but erects a wide and complex frame for possibilities of meaning. Ciardi says of metaphors and symbols, "They often manage to feel more than one way at once," and refers to them as "areas of refusal to be

¹Barfield, op. cit., p. 112.

²Bate, op. cit., p. 362.

specific in order to emphasize some larger possibilities of meaning."¹

Metaphor is one of the principal means by which, in Richards' words, a poem becomes a "set of possibilities," with dynamic, fluid meanings. In Richards' view, it is the activity of the reader in making the kind of connections demanded by metaphor that constitutes the real meaning of poetry. And Perrine, commenting on the power of metaphor to provide delight by involving the reader, says,

First, figurative language affords us imaginative pleasure. Imagination, in one sense, might be described as that faculty in the ability of the mind that proceeds by sudden leaps from one point to another. . . . The mind takes delight in these sudden leaps, in seeing likenesses between unlike things.²

But these very advantages in the use of metaphor, as in the case of ambiguity, have concomitant dangers for the inexperienced reader. Metaphor is perhaps the clearest illustration of poetry's typical tendency to "all-but-say" things. If skill in metaphor lies in not-quite-saying things, the reader, forced to make connections which are often extremely subtle and complex, can easily misread a poem. As in the case of ambiguity, a high degree of sensitivity to the context of the whole poem is necessary in the reader's adequate re-creation of metaphoric meaning. Otherwise he is likely either to overlook the metaphoric

¹Ciardi, op. cit., p. 711.

²Perrine, op. cit., pp. 58-60.

relationship or, as Richards found, to impose his own pre-suppositions on the relationship. As Turbayne says, the price of the use of metaphor is constant vigilance, the danger being of reading elements into the metaphoric relationship which violate the total impact of the poem.¹

"The victim of metaphor," says Turbayne, "accepts one way of sorting, or bundling, or allocating the facts as the only way."² Again, the answer appears to lie in Coleridge's exhortation to give "distinct and perpetual" attention to all the parts of a poem.

It is not unexpected, then, that students need help in learning to respond sensitively and observantly to metaphoric language. They need to be taught not to expect to exhaust the full meaning of any given metaphor or symbol and to avoid the temptations of symbol-hunting and reading-in meanings unwarranted by the context.

Sound. Pope's famous line in "An Essay on Criticism", "The sound must be an echo to the sense,"³ remains a remarkably accurate description of the part the sound of words and groups of words plays in the total meaning of a poem. It is important to note that "sound" is not an equal partner with "sense" in creating poetic

¹C. M. Turbayne, The Myth of Metaphor (New Haven: Yale University Press, 1962), p. 27.

²Ibid.

³Pope, op. cit., l.365.

meaning; it is, rather, an "echo". In other words, sound, dissociated from the denotative and connotative meanings of poetic language, has little or no value. On the other hand, if the poet or reader is not sensitive to the sounds of individual words and to the rhythms of phrases and sentences, the total poetic effect is significantly diminished, and the full value of a poetic experience cannot be communicated. Schiller, commenting on Richards' theory of poetry, points out that in Richards' view the source of the musical delight of poetry does not lie "simply in the pleasures provided by the sound of the verse."¹ However, as he goes on to say,

When we view the words of poetry as being invested with their meanings, sound patterns become a movement of meanings, not simply of sounds. It is this movement in which we delight. Using Coleridgean hints Richards thus assimilates the musical appeal of verse to the growth in meaning characteristic of the coadunative imagination.²

As many critics (Phenix,³ Barfield,⁴ Ciardi,⁵ Spender⁶) point out, in poetry, unlike music, sound is always intimately connected with other dimensions of poetic meaning. Words cannot escape being sounds produced by muscular action; neither can they escape being vehicles for

¹Schiller, op. cit., p. 27.

²Ibid.

³Phenix, op. cit., p. 182.

⁴Barfield, op. cit., p. 47.

⁵Ciardi, op. cit., p. 994.

⁶Spender, op. cit., p. 60.

other kinds of meanings. Ullman notes,

Where a sound happens to occur with a meaning to which it is naturally attuned, it will . . . add its own expressive force to the sense by a kind of "resonance" effect,¹

and the comments of Ogden and Richards, cited earlier, about the interdependent relationship between the sounds and associative values of words, reflect a similar point of view. It can be readily seen that most of the above-quoted comments are essentially footnotes to Pope's famous statement.

The specific ways in which sound contributes to poetic meaning can perhaps be most usefully examined by using the concepts of mimesis, onomatopoeia, and rhythm. Concerning the first two, Ciardi says,

The act of producing a word involves breath and muscle, and various kinds of muscular activity tend to produce various kinds of feeling. . . . When the muscular play tends to enact the denotations of the word as in "prestidigitation" or "oily" . . . then the word may be called mimetic. When the sound of a word imitates the sound of what the word denotes, (buzz, crunch) then the word may be called onomatopoetic.²

Ullman makes this same distinction, referring to "mimesis" as "secondary onomatopoeia."³

With regard to rhythm, Ciardi notes that

All the rhythms of poetry achieve their effect by the way they play against one another . . . [and against

¹Ullman, op. cit., p. 87.

²Ciardi, op. cit., p. 764.

³Ullman, op. cit., pp. 24-86.

the other elements of a poem, such as] diction, imagery, rhyme, line length, vowel qualities, consonant sequence, and grammatical structure. . . .¹

Thus rhythm, apart from acting in conjunction with other sound values, acts in conjunction with the other dimensions of poetic meaning. It should also be noted that the creating of rhythmical meaning is not simply a matter of strict adherence to a formal metrical pattern. As Walsh points out,

There is nearly always a tension between the theoretical rhythm of poetry and the way the words would be naturally spoken aloud if the poem were read like prose. This tension - one might almost call it a lover's quarrel . . . [exists] between the theoretical and actual rhythm.²

Because, as Murphy points out, most readers are accustomed to responding to words only referentially, they find it difficult to respond to the sounds of words, and need constant practice 'fooling around' with words to free themselves from what they take for granted and to begin to see--and hear--words as sources of delight in themselves.³

THE TESTING OF LITERATURE

Bruner, stressing the need for tests which will encourage and measure intuitive thinking, says,

¹Ciardi, op. cit., p. 994.

²Walsh, op. cit., p. 72.

³Murphy, op. cit., p. 70.

Usually intuitive thinking rests on familiarity with the domain of knowledge involved and with its structure, which makes it possible for the thinker to leap about, skipping steps and employing short cuts in a manner that requires a later rechecking of conclusions by analytic means.¹

Although Bruner is not referring specifically to literature or to the testing of poetry, his comments on intuitive thinking are eminently applicable to the evaluation of literary response, especially of response to poetry. His conclusion is similarly pertinent:

It is certainly clear that procedures or instruments are needed to characterize and measure intuitive thinking, and that the development of such instruments should be pursued vigorously.²

Authorities on English teaching are almost unanimous in stating that what matters most in the teaching of literature, something usually termed "appreciation", is extremely difficult and perhaps impossible to evaluate adequately. The difficulty is due, as Loban points out, to the inherent complexity of literature, which "is as vast and as deep as life itself,"³ and to the fact that student response to literature, as Loban, Ryan, Squire,⁴ and many others note, involves much more than factual recall and comprehension.

¹Jerome S. Bruner, The Process of Education (Cambridge, Mass.: Harvard University Press, 1960), p. 58.

²Bruner, op. cit., p. 61.

³Walter Loban, "Evaluating Growth in the Study of Literature," English Journal, XXXIV (November, 1947), 223.

⁴Walter Loban, Margaret Ryan, and James Squire, Teaching Language and Literature (New York: Harcourt, Brace, 1961), p. 309.

What it does involve appears closely related to Bruner's "intuition".

Many authorities also deplore the shortage of adequate measuring instruments and either directly or by implication offer certain suggestions as to the form the necessary instruments should take.

1. There is need for greater precision, i.e., knowledge of what is measured in each item, so that analysis of performance on individual items will produce more detailed knowledge about student response.

2. Research is dependent on the development of better (more valid and reliable) objective tests. There is a need for better objective tests. Although there is substantial agreement among English teachers about the virtues of the subjective, "open-ended" testing of literature, many authorities (Broening,¹ Budd,² Gunderson³) point to the relatively greater degree of precision and reliability provided by objective test items. The kind of comparison required for research, they argue, requires this kind of

¹Angela Broening, "The Development of Taste in Literature in the High School," in N. B. Smith (ed.), The Development of Taste in Literature (Champaign, Ill.: N.C.T.E., 1963), p. 48.

²Wm. C. Budd, "Research Designs of Potential Value in Investigating Problems in English," in Research in the Teaching of English (Champaign, Ill.: N.C.T.E., Spring, 1967), p. 8.

³Doris Gunderson, "Flaws in Research Design," in Research in the Teaching of English (Champaign, Ill.: N.C.T.E., Spring, 1967), p. 14.

reliability and precision. What is needed, they say, is more work in the development of new kinds of tests.¹

3. As Bruner and others repeatedly point out, there is a need for testing procedures which measure the student's knowledge of the fundamental structure of a discipline and which measure his ability to apply its basic principles. Too many literature tests continue to confine themselves to purely factual knowledge.

4. Closely related to an increased emphasis on structure is the use of works or passages previously unknown by the student to measure his grasp of basic concepts and principles rather than what he remembers of what he has studied in class.

5. There is a need for tests which will encourage and measure intuitive thinking, much of which is often falsely termed "guessing", so as to make "explicit provision for the cultivation of courage in taste."²

6. There is a need for the development of procedures for investigating a student's thought processes as he reads or solves a problem or responds to a test question. This is obviously very difficult because the activity is internal. It is therefore essential to devise measuring techniques which promote this "thinking aloud" and which permit it to be examined.

¹Budd, loc. cit.

²Bruner, op. cit., p. 67.

7. The use of variant, or "spoiled" versions appears to be a promising procedure. Many authorities in the teaching of literature, including Murphy¹ and Richards, point to the possible insights into the nature of student response provided by a technique of requiring students to choose between original versions of literary works and versions which have been deliberately distorted in various directions.

The above considerations serve as useful guidelines in shaping the design of the present study and in constructing the Poetry Test, as reported in Chapter III.

STUDIES OF STUDENT RESPONSE TO LITERATURE

In the last decade there has been increasing interest in the investigation of student response to literature, particularly to novels and short stories. Squire's study, The Responses of Adolescents to Four Short Stories,² which forms part of a series published by the National Council of Teachers of English, classifies responses under six headings: Literary Judgment, Interpretational, Narrational, Associational, Self Involvement, and Prescriptive Judgment. Squire's primary finding, that a major portion of the response (43 per cent) came under the heading

¹Murphy, op. cit., p. 102.

²James R. Squire, The Responses of Adolescents While Reading Four Short Stories (Champaign, Ill.: N.C.T.E., 1964).

Interpretational, is at first sight encouraging as a reflection of the over-all effectiveness of literature teaching. However, it is the writer's view that this finding was mainly a result of Squire's method of investigation, in which students reacted, orally, to a series of fragments of short stories. Squire further found that the quality of individual interpretations was generally unrelated to the intelligence and reading ability of the subjects as measured by standardized intelligence and reading tests.¹ He concluded that his findings concerning responses to short stories cannot safely be extrapolated to knowledge of student responses to other kinds of literary selections, and that additional research in response to other genres is needed. His main conclusion for the present study is that adolescent readers clearly need guidance in learning to interpret literature.²

Wilson, using Squire's scheme for classifying responses ("interpretational", etc.) studied responses of college freshmen before and after the study of selected novels. He found that the study of these novels increased the percentage of interpretive responses, and that there were marked individual differences in response to intensive study: "With some subjects interpretive adequacy increased

¹Squire, op. cit., p. 50.

²Ibid.

as the intensity of involvement decreased, [while with others] interpretive adequacy decreased as involvement decreased."¹ He, like Squire, concluded that a high degree of interpretational adequacy was rare.²

Purves, in a study published in 1968, attempted to categorize "The Elements of Writing about a Literary Work" in order to provide a framework for further empirical study. Though the kind of research undertaken by Purves does not lend itself to empirical conclusions, he raises a basic question touched on earlier in this chapter:

Does [self-conscious analysis] inhibit enjoyment of literature? After all, one may argue that engagement is the primary goal in literary education. . . . Without engagement - unspoken or spoken - there is little point in reading literature. It becomes a mechanical exercise. . . . To the extent that this is true, equally true is the counter that omnivorous reading without any thought is of tarnished value. The unthinking absorption of books is peripherally educational in the same way as is intricate analysis pursued for its own sake. There is, however, a balance, and this balance would seem to be the aspiration of education in literature. The teacher's function is to strengthen [the students' capacity to respond].³

It is precisely this "capacity to respond" with which the present study is concerned. Squire, Wilson and Purves, have concluded that high school students' response to literature requires careful cultivation. This cultivation

¹James R. Wilson, Responses of College Freshmen to Three Novels (Champaign, Ill.: N.C.T.C., 1966), p. 41.

²Ibid.

³A. C. Purves, Elements of Writing about a Literary Work (Champaign, Ill.: N.C.T.C., 1968), p. 63.

obviously demands exact knowledge of all aspects of this response. In attempting an analysis of student response to poetic language, this study seeks to provide more detailed information about response to literary language at the word (and phrase) level than has been supplied by these previous studies.

POETRY TESTS AND STUDIES OF RESPONSE TO POETRY

In 1923 Abbott and Trabue developed a test for measuring students' ability to judge between original and altered versions of whole poems.¹ This piece of research is reviewed here in some detail because of its uniqueness and because it, more than any other, provided a basis for the experimental test devised for the present study. The poems comprising the first items were arranged in an ascending order of difficulty, and the variants were devised so as to weaken the original "by diminishing or annulling one of its characteristic merits."² Accordingly, among these "characteristic merits" systematically distorted or weakened were "the emotional tone, the imaginative quality of the thought, and the rhythmic form."³ Thus, each item consisted of a choice among an original version, a

¹Allan Abbott and M. R. Trabue, A Measure of Ability to Judge Poetry (New York: Teachers College, Columbia University, 1925).

²Abbott and Trabue, op. cit., p. 4.

³Ibid.

sentimentalized version, a version in which the imagery was weakened or distorted, and a version in which the rhythm was interfered with. The investigators were concerned about the gap between the goal of developing critical appreciation advocated by the "theorists" and actual teaching and testing practice, and deplored the fact that most tests purporting to be measures of "literary appreciation" were actually measures of the ability to repeat a body of information.¹

Abbott and Trabue believed that

If it were practicable to devise an objective test of independent critical judgment applied to specimens of the art previously unknown, light would be thrown on the degree of success attained in developing the power to appreciate. . . . Conceivably something might be learned of the general psychological response to the stimuli of the art in question.²

Their main finding was that in the group tested, ranging from ten-year-olds to graduate students, the level of taste was extremely low, and that the test had doubtful validity for anyone below college level. This led them to question whether or not poetry should be kept out of the high school altogether, as something "too bright or good for nature's daily food". Another finding was that the test appeared to have more value as a teaching device--as a basis for discussion--than as a testing instrument. Most significant for the purposes of the present study, however, was their conclusion that the test

¹Abbott and Trabue, op. cit., p. 3.

²Ibid.

may well be used with high school classes for the information they furnish the teacher regarding the tastes of the pupils. Such diagnostic study of the elements of poetry . . . will be very helpful to any teacher who cares to undertake it.¹

The test devised by Abbott and Trabue, however, has certain serious shortcomings as an instrument for investigating with any degree of precision the way students respond to the various uses of poetic language. It does not necessitate intensive scrutiny of individual words and expressions as they interact with the whole context of a poem; it does not measure with any degree of precision a student's awareness of the various dimensions of poetic meaning operating in a word or expression, and the extent to which they can be regarded as discrete for teaching and testing purposes. In not requiring students to match the part to the whole it fails to capitalize on a natural pleasure in puzzle solving.

More recently, chiefly under the influence of the work of Bloom and his associates,² makers of poetry tests have used multiple-choice items in an effort to measure the "higher" levels of cognition such as application and analysis. This kind of test, although it could be considered an improvement over previous tests measuring only factual recall or knowledge of nomenclature, tends to ignore the

¹Abbott and Trabue, op. cit., p. 15.

²Benjamin Bloom, et al., A Taxonomy of Educational Objectives: Part I (New York: Longmans, Green, 1956).

distinctive operation of poetic language--as Ciardi puts it, "The way a poem performs itself."¹ Hence it appears that such tests may fail to measure the essential aspects of student response to poetry.

Probably the most ambitious, scholarly, and influential study of student response to poetry was conducted by Richards, who reported and discussed his findings in Practical Criticism, published in 1929. Richards, concerned about what he termed a general "levelling down" of literary taste in modern society, analyzed the written evaluations of thirteen poems by a large number of English literature students at Cambridge University. In his analysis he categorized the main reasons for failure to respond adequately to a poem under such headings as "stock responses", "doctrinal adhesions", "technical presuppositions", and "failure to make out the plain sense of a poem".² He found that few of his subjects, even though they were seniors specializing in English literature, could cope adequately with the poems. He found his "protocols", as he termed the students' written evaluations of poems, ridden with the grossest kinds of misapprehension and marked by a general tendency to prefer inferior to superior poems. During the course of his investigation he came to the

¹Ciardi, op. cit., p. 669.

²Richards, Practical Criticism, Chapters 4 to 8.

conclusion that

the closest scrutiny of details is compatible with the fullest, fairest, and most discriminating appraisal of the whole. Indeed, the two inevitably go together. The sovereign formula in all reading is that we must pass from judgment of details to judgment of the whole. It is always rash and usually disastrous to reverse the process. . . . The final decision can only be made through a fuller, closer contact with the poem itself; . . . it is [detailed] observation we seek in criticism.¹

Richards, like Abbott and Trabue, found a general inability on the part of his respondents to deal adequately with poetic language and a corresponding "low level of taste." Although Richards' work offers an intensive, comprehensive, well-documented analysis of a highly skilled critic, an analysis which has had considerable effect on the teaching of literature, it is not sufficiently objective and quantifiable to admit of replication or to serve as a practicable model for subsequent systematic investigation of response to poetry. As Schiller notes, Richards himself puts great stress on the value of methodology, but the greatest contribution of Practical Criticism is to critical and educational theory. In its emphasis on the interaction between reader and poem, Richards' work has had a profound effect on the theory and practice of literature teaching.

What the present study attempts to do, then, is to examine in a relatively detailed manner what Abbott and Trabue examined generally, and to examine in a fairly

¹Richards, op. cit., p. 38.

objective manner some of the detailed matters Richards dealt with subjectively. In this way, this study can be considered an extension of these two studies. It seeks, in a limited way, to combine the objectivity of the former with the detailed analysis of the latter.

SUMMARY

In Chapter I a review was made of statements concerning the need for research in the nature of student response to poetry and the corresponding need for the development of more adequate instruments for measuring this response. There was seen to be a serious shortage of this kind of research, essential in bringing necessary knowledge to the classroom teacher.

The present chapter attempted to develop the basis for a rationale for assessing student response to poetry by reviewing statements by literary critics and educators concerning both the nature of poetic language and the nature of literature testing. Also reviewed were investigations of response to literature in general and poetry in particular.

The main emphasis in this review was on the nature of poetic language as it differs from language used for discursive purposes. The central concern in the evaluation of student response to poetry was seen to be the need for devising instruments which measure something more than factual recall or ordinary comprehension as measured by

most standardized reading tests and which, in testing poetry in its own terms, reflect the Coleridgean view that a poem is an organic unit, its content and form inseparable.

The design of the present study, therefore, and the nature of the Poetry Test which forms the central part of the study, are a direct product of a rationale based primarily on the concepts regarding the nature of poetic language which were reviewed in this chapter and to a lesser degree on an examination of significant studies of student response to poetic language. This design and the rationale upon which it is based are presented in the following chapter.

CHAPTER III

EXPERIMENTAL DESIGN

INTRODUCTION

The main purpose of Chapter III is to describe the construction of the Poetry Test on which the major part of the study has been based. Also described are the Reading Test which was used in conjunction with the Poetry Test, the interviewing procedures developed for the study, the sampling procedures employed, and the procedures used in the collection and treatment of data. The experimental design required the construction of the Poetry Test and the administration of this test, together with the Reading Test and follow-up interviews, to eight classes of Grade X and XII English students in North Vancouver, British Columbia. Test scores were subjected to statistical analysis and the interview results compiled.

CONSTRUCTION OF THE POETRY TEST

General Rationale

An attempt has been made in the previous chapter to establish a rationale upon which decisions relating to the design of this study, particularly as they concern the construction of the Poetry Test, can be made.

This rationale embraces three principles which have been derived from what has been said about the nature of poetic language. First, poetic language, because of its economy, its precision, and its many dimensions of meaning, must be read intensively. Second, because of the organic unity of a poem and the interdependence of its structural elements, the reader of poetic language must be much more highly sensitive to context than the reader of discursive language. Third, because poetry is an art form making its appeal through the senses and is characterized by "all-but saying things", it requires active involvement by the reader if its meanings are to be communicated.

From these principles were derived the following guidelines for a valid poetry test:

1. The basic concern in the teaching and testing of poetry is the students' ability to derive meaning from the totality of the poetic language.

2. A valid poetry test should involve active response to whole poems, or reasonably integral sections of poems, and should focus upon the relation of individual formal elements to the whole context of a poem.

3. A valid poetry test should test in terms of the formal elements which function in unison to produce poetic meaning; only indirectly should it be concerned with students' factual knowledge of authors and of literary terminology (knowledge about poems).

4. A valid poetry test is only partially concerned

with measuring the paraphrasable "thought" content of a poem, i.e., with the skill in comprehension measured by most standardized reading tests.

5. A valid poetry test should test poetic elements on their own terms: it should test skill in the application of poetic principles rather than the ability to verbalize about these principles.

6. A valid poetry test must involve intensive scrutiny of individual words and expressions as they interact with the total context of a poem.

7. A valid poetry test must take into account the complex nature of poetic meaning, where several dimensions interact, affecting each other in a number of ways, often in a single word or expression.

8. A valid poetry test should attempt to make provision for responses at various levels of rationalization. In other words, it should encourage the kind of "guessing" Bruner considers in his discussion of the importance of fostering intuition in students. Further, the test should attempt to measure the level of conscious awareness generally involved in the response to poetry by high school students, and to determine whether this level varies with different types of items.

The Poetry Test eventually developed for this project attempted to incorporate these criteria by requiring students to choose among original and variant versions of whole poems. A word or group of words was deleted from the

poem and students (using clues from the context) were asked to select the "best" answer from four alternatives. It was hoped that this procedure would provide for both flexibility and precision in the measurement of various dimensions of poetic meaning (tone, metre, imagery, etc.). In allowing for use of distractors having varying degrees of difference from the original, or "right" answer, the procedure also promised to make possible a high degree of control over item difficulty.

Development of Categories

The next problem was to develop a number of discrete categories sufficiently comprehensive to reflect accurately the main dimensions of poetic meaning. Standard works on poetry, such as those by Brooks and Warren,¹ Wellek and Warren,² Drew,³ Ciardi,⁴ and Walsh,⁵ discuss poetic meaning under many different classification schemes. Sample headings from these works are "Metrics", "Rhyme", "Imagery", "Denotation and Connotation", "Metaphor", "Personification",

¹Cleanth Brooks and R. P. Warren, Understanding Poetry (New York: Holt Rinehart, 1960).

²R. Wellek and Austin Warren, The Theory of Literature (New York: Harcourt, Brace, 1956).

³Elizabeth Drew, Directions in Modern Poetry (New York: Gordon Press, 1967).

⁴John Ciardi, How Does a Poem Mean (Boston: Houghton-Mifflin, 1959).

⁵Chad Walsh, Doors into Poetry (Englewood Cliffs: Prentice-Hall, 1962).

"Metonymy", "Symbolism", "Allegory", "Paradox", "Understatement", "Irony", "Tone", "Musical Devices", and "Pattern".

It was therefore necessary to develop a somewhat independent scheme cutting across existing classifications. Using Pope's famous line and Perrine's title¹ as a base, two "sound" and two "sense" categories were initially constructed: "rhythmical" and "non-rhythmical" sound, and "denotative" (plain sense) and "connotative" (poetic) meaning.

It was decided to use four rather than a larger number of categories, partly because it was recognized that such elements as "imagery" and "tone" rarely act independently of one another, and partly to make the interpretation of test results more manageable. Also because the testing procedure required an average of almost two minutes per item (each item involved the reading of a whole poem), it was found necessary during the pilot studies to limit each test form to 16 items. With four categories, this meant four items per category, which seemed a minimum for purposes of analysis.

Test items were constructed for each of the four categories and tried out on high school students and colleagues. It was discovered that constructing items for the "rhythmical", or metric, and for the "thought"

¹Laurence Perrine, Sound and Sense (New York: Harcourt, Brace, 1963).

categories presented no great problem. The "suggestion" and "sound" categories were another matter. In these categories responses failed to coincide with the intention of items to a sufficient degree to justify further attempts to modify the categorization scheme. Consequently, "suggestion", or "poetic meaning", was replaced by Walsh's term "indirection",¹ which includes such poetic elements as figurative language and irony. "Non-rhythmical sound" was replaced by "nuance",² also Walsh's term, which includes the connotations and sensuous aspects of words as well as their sound values. However, in constructing items for the revised categorization scheme, the difficulties encountered with the "nuance" and "indirection" categories were even greater than those encountered with the first scheme.

At this point it was decided that these problems in categorization were due primarily to the nature of poetic language itself--to the way in which the various dimensions of poetic meaning operating in a word act interdependently. This interdependence appeared to be particularly strong between, on the one hand, the evocative or sensuous quality of a word and, on the other, its sound and its metaphoric or symbolic sense. For example, if "Much have I traveled in the realms of gold" is changed to " . . . lands of gold," the "plain sense" and metre remain basically the same, but

¹Walsh, op. cit., p. 51

²Walsh, op. cit., p. 34.

to some degree the sound, and to a high degree the suggestive value of the phrase are altered. As Richards suggests, it is impossible to alter the sound of a word or phrase without substantially altering its connotative value because the sounds of individual words and expressions act only as a reinforcement to their suggestive qualities.¹ A similar example could be given to illustrate the close relation between connotation and metaphor which makes it impossible to alter the connotative value of a word without altering its metaphoric effect. Contrast "Like a rich jewel in an Ethiope's ear" with "Like a rich jewel in a woman's ear": all the power generated by the light-on-dark imagery is dissipated.

Consequently, with some reluctance, it was decided to combine connotation, imagery, metaphor, and word sound, into a single category, to be designated "poetic meaning". Because this was considered by far the most important category, 50 per cent of the items were assigned to it, with 25 per cent assigned to "thought" and 25 per cent to "rhythm". However, during the final construction and selection of test items and the statistical analysis of test results, the original "sound" category was restored. This category, which consisted of items in which the sound of individual words and expressions contributes significantly

¹J. P. Schiller, I. A. Richards' Theory of Literature (New Haven: Yale University Press, 1969), p. 27.

to the total poetic meaning, was, after much deliberation, judged to be sufficiently distinct from the "suggestion" category to be treated as a separate unit. It was reasoned that the restoring of the "sound" category would permit a useful examination of the relationship between responses on the "sound" subtest and responses on the other three subtests. Thus, the original four categories, i.e., Thought, Suggestion, Sound, and Rhythm, were substantially the ones which were finally used.

The Selection of Poems to be Used as Test Items

The criteria for the selection of poems were:

1. Brevity. Only whole poems no longer than 18 lines and a few reasonably integral sections of longer poems were used.

2. Student Interest. In order to gain a maximum amount of student interest and candor in responding to the test, the pilot study results were used in an effort to select poems judged likely to be at a high interest level for most high school students.

3. Literary Quality. An attempt was made to include only poems of reasonably high literary quality. This was partially ensured by the fact that almost all of the poems were taken from recognized anthologies, many of them actually in use in high school and university courses.

4. Unfamiliarity. It was essential that the Poetry Test items not merely measure factual recall of words and

phrases from familiar poems. Thus, an attempt was made to exclude poems actually listed in senior high school curriculum guides for English courses for Alberta, where the pilot test was to be administered, and British Columbia, where the final test was to be administered.

5. Freedom from Unusual Allusion and Vocabulary.

In order that the test items would not be measuring factual recall, whether of historical or mythical material, or of the meanings of unfamiliar words, an attempt was made to exclude poems containing unduly difficult allusions and vocabulary items. In a few instances, word meanings were appended to items.

Construction of Test Items on the Basis of the Categories

As has been indicated, the development of test categories and of individual items was a reciprocal process, the categorization scheme providing guidelines for item construction and the item construction providing a check on the adequacy of the categorization scheme. From the more than 120 poems selected according to the criteria listed in the previous section, items were constructed for each of the four categories. Item construction in general proved relatively difficult, but items for Categories One and Four (Thought, and Rhythm) proved easier to construct than items for the other two categories (Suggestion and Sound).

Distractors for items in Category One were constructed so as to alter the "thought", or "plain sense meaning" of

the poem, while keeping constant the metre, and, as far as possible, the connotation and sound. Distractors for items in Category Four were constructed so as either to alter the nature of the metrical feet in a line or to vary the number of feet within a line. The construction of distractors for Categories Two and Three involved such changes as distorting or weakening imagery and changing the connotative and sound values of words. This problem proved exacting, and was, in the writer's judgment, not solved in an entirely satisfactory manner in the course of the study.

With the help of Roget's Thesaurus, and after much revision of individual distractors, 120 items, distributed among the four categories, were finally assembled.

Item Selection

From these 120 original items, six tests of 15 items each were administered over a period of three months, four to eight Grade XI classes in the City of Edmonton and the County of Strathcona, Alberta, and two to four Grade XII classes in North Vancouver, British Columbia. Item analysis was performed in order to determine measures of difficulty and discriminating ability for each item. These measures, together with a scrutiny of the response patterns for each distractor, made it possible to ascertain which items seemed sufficiently suitable to be retained intact, which items were to be rejected, and which items seemed to be reparable. For the purpose of these "pre-pilot" studies,

items with a difficulty index below 0.30 and above 0.70 and items with a biserial correlation of less than 0.35 were generally considered unsuitable. Similarly, items in which any one of the alternatives gained either more than 90 per cent or less than 5 per cent of the total response were judged to be in need of repair.

Standard works on test construction tend to hold that item difficulties (proportion of correct responses) should, except for special screening requirements, cluster around the 0.50 level. Because the results of item analysis on the "pre-pilot" tests showed a mean difficulty index of between 0.35 and 0.40 a general discarding of the more difficult items was necessary.

With regard to the level of discriminating power of an item, a figure of 0.70 or higher is considered satisfactory by most authorities for a test purporting to measure a homogeneous skill. Because the Poetry Test, as reflected in its categorization scheme, was concerned with measuring skills at least partially discrete, a considerably lower biserial correlation coefficient was deemed a satisfactorily high index of an item's discriminating power.

With respect to the distribution of student responses among the four alternatives offered in each item, an initial attempt was made to obtain a "50-17-17-17" pattern, the 50 representing the percentage of "correct" responses, and the incorrect responses distributed equally among the three distractors. It was later determined that more information

could be gained from such patterns as "50-5-30-15" and "70-15-10-5" where, say, in an item involving metre, a distinction could be made between students with some grasp of the operation of metre and students with virtually no grasp at all. Providing for differing response loadings on the various distractors, then, appeared to make it possible to separate grosser from less gross distinctions among the alternatives, and hence among students' degrees of skill.

Fifty-four items, distributed reasonably equally throughout the four categories, survived these various tests for item selection.

Validation by Judges

The remaining 54 items were then examined by four judges selected for their special competence in the area of English literature and literature teaching.

These authorities, all experienced readers and teachers of English poetry, were given a description of the test categories and asked to examine each item and to assign it to one of the four categories. The judges were also asked to comment on any weaknesses noted in any of the items and to identify any items which were considered very difficult to categorize.

In comparing the categorization by the judges of all 54 items with the intended category for each item it was found that for Judge Number One agreement was 96 per cent,

for Judge Number Two agreement was 81 per cent, and for Judge Number Three, agreement was 91 per cent. Because of the pressure of time, Judge Number Four was able to examine and categorize only 17 of the items; his rate of agreement was 97 per cent. Five items were discarded because of flaws detected by the judges, leaving 49 items, and a further nine items were discarded as a result of a re-examination of the items by the investigator for the purpose of creating two parallel test forms. This left 40 items, for which the mean rate of agreement between the judges and the experimenter and intended item categories was 94.8 per cent, and the mean inter-rater, or "inter-judge", rate of agreement in categorization was 95.7 per cent.

Construction of Parallel Forms

After the items questioned by one or more of the judges were discarded, 49 items remained. As was indicated in the previous section, two groups of 20 items each were then formed to serve as parallel test forms. An attempt was made to match the two forms in mean item difficulty level and in mean item discrimination level.

Equal numbers of items from each category were used in each of the forms and an attempt was made within categories to match items as closely as possible for difficulty and discriminating power. Although every effort was made through item selection and revision to create two test forms as nearly equivalent as possible, significant variation in

item analysis results from certain identical items administered to different groups in the "pre-pilot" studies indicated that statistical parallelism was not possible.

The Pilot Study

The two comparable forms labeled Form A and Form B were then administered to two Grade XI classes in the City of Edmonton, Alberta. Subsequently follow-up interviews were conducted with a random sub-sample of four students. The test results were then subjected to item analysis.

The results of the Pilot Study revealed no need to revise either the Directions to Students which prefaced the test (see Appendix A) or the mechanics of test administration. It was found that while most respondents required an average of approximately a minute and a half per item, a substantial minority required up to two minutes per item. Since the test was designed to serve as a power rather than a speed test, it was decided to allow two minutes per item in the final administration.

As a result of observed student reaction to the length of the test during the actual writing and opinions expressed during the interviews, the final test forms were constructed so as to consist of 16 rather than 20 items. Because of the demanding nature of the test many students appeared to lose interest and to be inclined to resort to guessing after completing approximately 15 items. Reducing the number of items from 20 to 16 had the added advantage

of making possible a further balancing of the two forms by allowing for the discarding of items which failed to match items in the opposite test form.

Final Revision of Test Forms

The item analysis performed on the results of the Pilot Test administration resulted in the above-mentioned discarding of eight items, two from each category, which failed to meet the three criteria listed earlier relating to the difficulty, discriminating power, and response pattern of each item. Also discarded were any items with distractors receiving less than five per cent of the total response. Also as a consequence of item analysis, certain items from Form A were exchanged for items in the same categories from Form B so as to produce greater equivalence in the Forms. The operations of discarding and rearranging resulted in both forms having a mean item difficulty index of 0.45 and a mean biserial correlation coefficient of 0.49.

Although the two test forms which emerged from the final Pilot Study appeared to be equivalent in mean item difficulty and mean item discriminating power, a comparison by categories revealed small differences for Categories One and Four (Thought and Rhythm). For example, the mean item difficulty index for Form A, "Thought", was 0.51, while for Form B, "Thought", it was 0.46. For Form A, "Rhythm", the mean item difficulty index was 0.41 while for Form B, "Rhythm", the mean item difficulty index was 0.49. Because

of the variation previously noted in performance on identical items by different groups, these differences were not judged sufficiently serious to warrant further revision of the test forms.

INTERVIEWING PROCEDURE

The following interviewing pattern was developed as a result of trial sessions held during the many "pre-pilot" studies and during the Pilot Study:

1. An initial period of at least five minutes was provided for establishing rapport between the interviewer and the respondents and for explaining the purposes of the study. During this period such questions were asked as "Did you have enough time?" and "How did you attack the questions?" and respondents were invited to give, at any time during the interview, advice for improving any part of the investigating procedure, including the interview itself. It was made clear that adverse criticism was welcome, and that the whole value of the study depended on honesty in student responses.

2. The tape recorder was not switched on until a certain degree of rapport was attained. Students were told that they could ask for it to be shut off at any time while they pondered any of the questions. This was necessary because the interviews were conducted between two and seven days after the administration of the Poetry Test.

3. In the "body" of the interview, the central question was "Why did you select that particular alternative?" To make possible some kind of comparative analysis of interview results, the body of the interview, taking about 40 minutes, consisted of a set questioning schedule (see Appendix C) where essentially the same questions were asked of each respondent, in the same order, for each item. In addition, to obtain information not obtained through the planned questions, and to provide some degree of flexibility, allowance was made for free response at any time. Students were repeatedly urged to comment freely at any time during the interview on any of the questions previously raised or on anything at all which appeared relevant.

4. A final short period following the body of the interview was provided in which students were again urged to comment freely on any relevant matters. Thus it was hoped to gain a balance between the relatively fixed core of the interviews and an area of free response surrounding this core.

Following the final administration of the Poetry Test, a representative sub-sample of ten students, selected by grade level, by sex, and by level of achievement in the test, was obtained for the interviews.

Because of the exploratory nature of the study, the importance of the interview results as a supplement to the

data provided by the Poetry Test was considered sufficient to warrant the reporting of the results of the interview in a separate chapter. The findings arising from the interviews will be presented in Chapter V.

THE READING TEST

The ability under investigation in the present study, skill in responding to poetic language, was regarded partly as a special kind of reading skill. Incorporated in the study, therefore, was the administration of a standardized reading test as one means of construct validation for the Poetry Test. The main point at issue was, how closely would Reading Test Scores correlate with Poetry Test Scores? Also, if it could be found that a reading test had been established by subsequent research as correlating highly with intelligence and academic achievement, it would be possible to examine further the construct under investigation, namely response to poetic language.

After consulting authorities in reading at the University of Alberta and the Mental Measurements Yearbook (6th) edited by Buros,¹ it was decided to use the Nelson-Denny Reading Test, Form A.² This is a test of well established reliability which could be administered in under 40 minutes and which measures skill in Vocabulary,

¹O. K. Buros, The Sixth Mental Measurements Yearbook (Highland Park, N. J.: Gryphon, 1965).

²M. J. Nelson and E. C. Denny, The Nelson-Denny Reading Test: Examiner's Manual (Boston: Houghton-Mifflin, 1960).

Comprehension, and Reading Rate. Further, as the authors point out:

Research . . . indicated a close relation between test scores and scholastic achievement. . . . In a summary of 57 reported correlations between scholastic achievement and achievement in test scores in specific subject matter fields, the range of coefficients was from .10 to .70, with a median of .48. The Nelson-Denny Test showed a correlation of .67.¹

Thus the Nelson-Denny test was judged to be at least as useful for the purposes of this study as any other of the many commercially available reading tests.

It was decided, therefore, to administer the Nelson-Denny Reading Test in addition to the Poetry Test. Because each test requires slightly more than a half hour, and because it was considered desirable to minimize test fatigue, the tests were administered to each class in two consecutive periods.

SAMPLING PROCEDURE

The sample of pupils involved in this study was drawn from a population of about 2400 pupils consisting of all Grade X and Grade XII pupils taking the academic English programme in North Vancouver, British Columbia. Of the six Senior Secondary Schools in North Vancouver, only five were eligible because one had participated in one of the pre-pilot studies. From these five, two schools were selected at random, and from each of these two schools two

¹Nelson and Denny, op. cit., p. 26.

Grade X and two Grade XII classes were randomly selected. The result was a sample of 191 pupils in eight classes, four classes from Grade X and four classes from Grade XII. Since the Poetry Test and the Reading Test were administered on different days, some students writing one of the tests failed to write the other. The Reading Test results from pupils who failed to write the Poetry Test were discarded, and the Poetry Test results from pupils who failed to write the Reading Test were not used where findings required a comparison between the two tests. Thus the sample size for statistical procedures involving only the Poetry Test is larger than that for procedures involving both tests. The total sample sizes, respectively, were 191 and 165. For students writing Form A, the respective figures are 93 and 85; for Form B, 98 and 80.

The sub-sample used for the interviews was selected so as to represent students achieving low, medium, and high scores on the Poetry Test. Comprising a total of ten students, it included two Grade XII students who achieved the two highest scores in the total sample; one Grade XII student and three Grade X students selected randomly from scorers above the 80th percentile in each grade group; one Grade XII student selected randomly from scores near the 50th percentile; and one Grade XII and two Grade X students selected randomly from scores between the 20th and 40th percentile in each grade group.

DATA COLLECTION

The data collected for each subject consisted of Grade Level, Sex, Poetry Test Totals and Subtotals, and Reading Test Totals and Subtotals. The Poetry Test and Reading Test, each requiring approximately half an hour, were administered on two separate occasions, each student answering only one form of the Poetry Test because of time limitations.

In addition, the ten students were interviewed from two to seven days after the writing of the Poetry Test, and the results recorded on tapes.

STATISTICAL TREATMENT OF THE DATA

Item Analysis. Item analysis for the test as a whole was conducted for all items in each of the two Test Forms in order to determine the difficulty level index, an index of item discriminating power, and the response pattern operating within the alternatives for each item. The index of item discriminating power was the biserial correlation coefficient. Since this coefficient is difficult to interpret in a heterogeneous test such as the Poetry Test, a separate item analysis was also conducted within each of the four subtests. This procedure had the effect of substantially raising the numerical value of the biserial correlation coefficients, but it so reduced the number of items involved that it was considered statistically unsound.

However, because it did serve to aid somewhat in interpreting the biserial correlation coefficients, the results of the item analysis within subtests are included in Appendix F.

Intercorrelations, means, and standard deviations.

Means, standard deviations, and intercorrelations were obtained for the eight scores comprising the Poetry Test totals and subtests and Reading Test totals and subtests. The purpose of obtaining these intercorrelations was to examine possible relationships between performance on the Reading Test, the two forms of the Poetry Test, and the various subtests of both the Reading and Poetry Tests.

Analysis of variance. Several two-way analyses of variance were conducted in order to compare relationships between Test Form, Sex, Grade Level, Reading Test Scores and performance on the Poetry Test totals and on each of the subtests. Students were divided into top, middle, and low groups according to reading test scores for this purpose.

Factor analysis. A principal components analysis was performed on the results from each of the Poetry Test Forms in an attempt to determine the nature of the underlying factor structure in response to poetic language.

SUMMARY

This chapter presents a general description of the design of the study and the various procedures used in the

investigation. The main emphasis is on the procedures followed in the construction of the Poetry Test. Also described are the development of a suitable interview procedure, selection of the Reading Test, sampling procedures, data collection, and treatment of the data.

The findings emerging from analyses of the data resulting from the Poetry Tests and Reading Tests are reported in Chapter IV. The findings emerging from the recorded interviews are reported in Chapter V.

CHAPTER IV

RESULTS AND DISCUSSION

INTRODUCTION

Before presenting and discussing the results of the statistical procedures used in the analysis, some brief comments on the general reporting design seem in order. First, the findings of the study are presented under four main headings: Poetry Test Characteristics, Performance on the Poetry Test by Various Groups, Factors in Response to Poetic Language, and Interview Results. Because the last of these is non-statistical in nature and supplementary to each of the other three, the results of the interviews will be presented in a separate chapter.

Second, it is obvious that in getting and analyzing responses to test items, an investigator obtains information about both respondents and the items themselves. Thus a clear line cannot be drawn between statistical procedures which lead to information about tests and test items and those which lead to information about the performance of individuals on a particular test or item. Nevertheless, in any given procedure there tends to be a focus. For example, in item analysis the focus is clearly on the test or test item itself, while in analysis of variance with the test

scores as the criterion measure, the focus is on differences in performance among various groups of individuals. The reporting of the statistical findings of this study is therefore organized in accordance with the main focus of each procedure.

Third, it was decided to incorporate the interpretation and discussion of each of the kinds of data with the presentation of the data rather than to present the discussion subsequently as a whole. While this arrangement interferes to some degree with the continuity of the discussion, it presents fewer problems to the reader in relating the discussion to the data.

POETRY TEST CHARACTERISTICS

Reliability

A central consideration in the construction and use of any test is reliability, or "the accuracy with which a score represents the status of an individual in whatever aspect the test measures him."¹ Because a test lacking this accuracy cannot be valid, it is essential to obtain information about its reliability.

The reliability of a set of measures is theoretically defined as the proportion of true variance, as opposed to error variance, in the total variance. Operationally,

¹J. P. Guilford, Psychological Measurement (New York: McGraw-Hill, 1954), p. 349.

reliability has several related but separate meanings, each associated with the type of procedure used in its estimation. Authorities in the field of measurement note that the type of estimating procedure most appropriate for a given set of measures depends on the purposes for which these measures are to be used and on practical considerations such as availability of data in the proper form, testing conditions, and the kind of test used.¹

Three main types of procedures for estimating reliability have to do with (1) internal consistency, (2) alternate forms (or parallel forms) reliability, and (3) retest (test--retest) reliability. With the present test, a measure of internal consistency as well as of stability was considered desirable because the test is composed of four discrete sub-tests, each designed to be homogeneous with itself. For this reason the test--retest procedure, which provides no information about internal consistency, was rejected. The alternate-form method, which can provide information about both stability and internal consistency, was considered as the most appropriate procedure, but since design limitations precluded the administration of both test forms to the same subjects the intercorrelation between the two forms could not be obtained (see Chapter III, p. 86). Thus the alternate-forms procedure was rejected in

¹J. P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill, 1965), p. 446.

favour of a purely "internal consistency" estimate, involving the use of the K-R 20 formula. The chief weakness of this procedure in the present case is that it provides no information about test stability. This deficiency is partly remedied by the interview data reported in Chapter V, which yield some information about the stability of the test. Because the present test may be considered a highly "reactive" measure there is a danger of obtaining a spuriously high reliability coefficient through use of test-retest and equivalent forms methods. The use of the K-R 20 formula, which is purely a measure of internal consistency, was thought to minimize this danger.

Table I presents K-R 20 reliability coefficients for the two test forms for the total sample and the two grade level groups separately. In drawing inferences from these coefficients certain points should be noted:

TABLE I
K-R Reliability Coefficients for Forms A and B

Source of Scores	Grade X	Grade XII	Total Sample
Form A	.21	.35	.36
Form B	.26	.64	.58

1. It is generally recognized that the K-R 20 estimate of reliability tends to underestimate the

reliability of a test.¹ One of the chief reasons for this is that the use of the K-R 20 formula depends on conditions rarely met in achievement tests, and only partially met in the present test. As Guilford notes, "The derivation of the K-R formula rests on assumptions of a unifactor test. . . ."2 He also notes that with short tests (9 to 18 items) such as the present one, there is a pronounced negative bias.

2. The present test was designed to be a multi-factor rather than a uni-factor test and was generally successful in this regard, as indicated by the low correlation between sub-test totals and by the presence of several factors (reported later in this chapter); therefore the low overall reliability coefficients are not unexpected. In view of the complexity of the skill under investigation (see Chapter II) these figures, particularly those for Grade XII in Form B, appear to reflect a reasonable degree of internal consistency.

3. It will be seen from Table I that the reliability coefficients are considerably higher for the Grade XII group than for the Grade X group. One explanation for this difference is that the K-R 20 formula assumes approximately equal item difficulties. An examination of Tables XIV (p. 111) and Table XV (p. 114) reveals substantially greater consistency in item difficulty indices for the Grade XII

¹Guilford, Fundamental Statistics, p. 461.

²Guilford, Psychological Measurement, p. 383.

subjects than for the Grade X subjects. Another explanation is that because the test tends to be too difficult for the Grade X students, the Grade X scores are generally closer to the chance level than the Grade XII scores, reducing the item-total score intercorrelations (see Tables XIV and XV). The data presented in Table I suggest that neither test form is sufficiently reliable for Grade X students to serve as a basis for valid interpretation.

4. The data presented in Table I also show higher K-R 20 coefficients for Form B than for Form A, especially with the Grade XII group. This difference between test forms is reflected in the mean item data figures presented in Table II, where again the relatively greater degree of internal consistency for Form B is more marked for the Grade XII group than for the Grade X group.

TABLE II
Comparison of Mean Item Data Figures for Forms A and B

	Grade X		Grade XII	
	Form A	Form B	Form A	Form B
Mean Difficulty Indices	.33	.39	.44	.51
Mean Biserial Correlations	.39	.39	.41	.53
Mean Item Rel. Indices	.17	.17	.19	.25

Perhaps the most that can be said by way of explaining the difference between the K-R 20 coefficients

for the two test forms is that in spite of efforts at matching, Form B items on the whole seem to either possess greater discriminatory power or to measure more idiosyncratic abilities than Form A items.

As a final comment regarding the reliability of the test as a whole, two points can be noted:

1. In an exploratory investigation such as the present one, the main purpose is to extract a maximum amount of information about both the possibilities of the test and the responses of the students rather than to achieve a high reliability coefficient. Naturally, with the reliability of a test sufficiently low, little information is generated. In view of the complexity of the skill under investigation and the novelty of the investigating technique, the reliability coefficients presented in Table I appear reasonably satisfactory, at least for the Grade XII scores.

2. Guilford points out that while it is important for achievement tests and tests used in clinical diagnosis and personnel classification to be highly reliable, for research purposes, much lower reliability can be tolerated. Guilford further states,

If we are attempting to predict criteria that are very complex factorially, the best way to improve the test validity is to bring in additional common factors that are also in the criterion, thus increasing validity by lowering internal consistency reliability.¹

¹Guilford, op. cit., p. 389.

3. It is well known that reliability can be increased by increasing the number of items in a test. Because the present test is necessarily short (16 items), and because it proved impractical to administer both forms to the same students, it seems reasonable to estimate the probable effect of doubling the test length on reliability by means of the Spearman-Brown formula.¹ As Table III indicates, such a doubling of test length produces quite respectable reliability coefficients, particularly for Form B and for the Grade XII scores.

TABLE III

K-R 20 Coefficients Estimated for a Test of 32 Items
Using the Spearman-Brown Prophecy Formula

	Total	Grade X	Grade XII
Form A	.53	.35	.52
Form B	.74	.41	.78

Comparison of the Test Forms as a Whole

Means and standard deviations. It will be seen from Table IV that both means and standard deviations are considerably higher for Form B than for Form A. However, the obtained t-values which are presented in Table V show that,

¹Guilford, op. cit., p. 353.

TABLE IV

Comparison of Means and Standard Deviations for
Poetry Test Totals for Forms A and B

	Total Group		Grade X		Grade XII	
	Form A	Form B	Form A	Form B	Form A	Form B
Mean	6.27	7.00	5.47	5.78	7.16	8.22
SD	2.25	2.78	1.87	2.95	2.31	2.87

TABLE V

Significance of Difference of
Form A Means and Form B Means

Source of Scores	Obtained	t-values	
		p = 0.05	p = 0.01
Total Sample	1.98*	1.96	2.58
Grade X	0.80	2.02	2.70
Grade XII	1.93	2.02	2.70

*p < 0.05

while there is a significant difference between the total group means of Forms A and B at the 0.05 level of significance, no significant difference exists between the means for the two grade groups.

Thus the differences between means and standard deviations for the two test forms appear sufficiently small to warrant considering the two forms comparable. The forms, however, cannot be considered parallel since for this condition to be met the means and standard deviations must be equal.

An examination of Table IV suggests that both Forms are perhaps too difficult for Grade X students. The Grade X means are sufficiently close to the chance level--a score of approximately 4 out of 16 on the present test--to cast some doubts on the validity of both Poetry Test Forms for the Grade X level.

Mean item difficulty indices and biserial correlation coefficients. The mean item difficulty for both grade levels and for the total sample as presented in Table VI is substantially higher for Form B than for Form A. This is consistent with the differences in test means presented in Table IV and suggests not only that both Forms are more suitable for Grade XII students but also that even for Grade XII students, only Form B approximated the intended mean item difficulty level of 0.50.

TABLE VI
Comparison of Mean Difficulty Indices and
Biserial Correlations for Forms A and B

	Difficulty		Biserial Correl.	
	Form A	Form B	Form A	Form B
Grade X	.32	.38	.39	.39
Grade XII	.44	.51	.41	.53
Total	.38	.45	.40	.46

The mean biserial correlation coefficients shown in Table VI indicate that the discriminating power of the two Forms is similar, but that here again Form B appears to be somewhat superior.

Biserial correlation coefficients for individual items indicate the degree of homogeneity of a given test as well as the discriminating power of a given item. Because the skill the Poetry Test attempts to measure is perhaps more heterogeneous than homogeneous in nature, these coefficients must be interpreted with care. It is difficult to determine, for example, to what extent a low biserial correlation coefficient for a given item indicates a low level of discriminating power and to what extent it indicates that the item is measuring a skill which is different from that measured by the other test items. Therefore, as a supplement to the main item analysis, each of the subtests in each form was treated as a separate test and a second

item analysis carried out. The results are presented in Table VII. This procedure resulted in mean biserial correlation coefficients of 0.65 for Form A and 0.69 for Form B. While the small number of variables involved (in effect each test consisted of only four items) made it impossible to compare these mean coefficients with any degree of statistical validity¹ their similarity suggests that the two test forms are similar in overall discriminating ability.

TABLE VII
Mean Biserial Correlation Coefficients
Using Subtest Totals

Subtest	Mean Biserial Correlations for the Four Subtest Items	
	Form A	Form B
"Thought"	.65	.66
"Suggestion"	.49	.67
"Sound"	.74	.73
"Rhythm"	.72	.79
Total Means	.60	.71

Intercorrelations between total scores and subtest scores. Table VIII presents a comparison of intercorrelations between subtest totals and test totals for the two

¹Quinn McNemar, Psychological Statistics (New York: Wiley, 1962), pp. 139-40.

TABLE VIII

Comparison of Intercorrelations between
Poetry Test Total Scores and
Subtest Scores for Forms A and B

Subtest	Subtest Score with Total Score	
	Form A	Form B
1. <u>Thought</u>		
Total Group	.63	.67
Grade X	.70	.55
Grade XII	.49	.76
2. <u>Suggestion</u>		
Total Group	.66	.59
Grade X	.58	.33
Grade XII	.71	.64
3. <u>Sound</u>		
Total Group	.49	.74
Grade X	.32	.56
Grade XII	.62	.81
4. <u>Rhythm</u>		
Total Group	.60	.63
Grade X	.51	.62
Grade XII	.56	.60

test forms and for the two grade groups. Generally, these intercorrelations are moderately positive, and are similar for the two test forms, tending to be slightly higher for Form B than for Form A. The presence of markedly lower intercorrelations between the "Sound" and "Suggestion" totals and test totals for the Grade X group perhaps reflects the relative inability of the Grade X students to deal with the items in these two categories--an inability which is suggested by results reported elsewhere in this chapter.

Total scores and reading test scores. The correlations of the total scores on Form A and Form B with the Reading Test scores for the total sample, 0.48 and 0.44 respectively, are presented in Table IX. The similarity between these coefficients again suggests that the test forms are reasonably comparable. However, the difference between these coefficients was tested using Fisher's z-transformation technique¹ and was found to be significant at the .05 level. This, together with the fact that all the coefficients reported in Table IX are significantly greater than zero, suggests that the Poetry Test and the Reading Test measure discrete but related abilities.

¹J. P. Guilford, Fundamental Statistics, pp. 189-90.

TABLE IX
Correlations between Poetry Test Scores
and Reading Test Scores

Reading Test Scores	Poetry Test Scores	
	Form A	Form B
Total	.48*	.44*
Vocabulary subtest	.42*	.46*
Comprehension subtest	.46*	.31*

*p < 0.01

A comparison of increases in Poetry Test and Reading Test performance of Grade XII students over Grade X students is presented in Table X. The per cent increases for the two forms of the Poetry Test and for the Reading Test indicate that overall, there is a slightly greater increase in Poetry Test performance than in Reading Test performance. However, for Form B, which appears on the basis of the previously reported data to be the more reliable test form for Grade XII students, the Poetry Test increases are substantially greater than the Reading Test increases. This may provide a further indication that the two tests measure somewhat different abilities, and possibly suggests that during the high school years there is a greater gain in students' responsiveness to the linguistic subtleties embodied in poetic language than in the students' responsiveness to the "normal", or referential, operation of language.

TABLE X

Comparison of Gain in Performance of
Grade XII Over Grade X Students on
the Poetry Test and the Reading Test

Source of Scores	Grade X Mean	Grade XII Mean	Per cent Increase
<u>Poetry Test</u>			
Form A Group	5.5	7.2	31
Form B Group	5.8	8.2	42
<u>Reading Test</u>			
Form A Group	61.9	87.4	41
Form B Group	65.0	81.7	26

Characteristics of the Four Subtests

Means and standard deviations of subtest scores.

Table XI shows that generally there is a fairly high degree of similarity between standard deviations both as compared among the four subtests in each Test Form and as compared between corresponding subtests of Forms A and B. This similarity holds true for the total sample and for both grade groups, suggesting that the subtests produced fairly similar dispersions of scores, thus permitting a comparison of the mean scores of the four subtests.

A comparison by inspection of the subtest means of the two test forms reveals somewhat greater differences

TABLE XI

Comparison of Means and Standard Deviations
of Subtest Scores

TOTAL SAMPLE								
Test Form	Subtest 1 Thought		Subtest 2 Suggestion		Subtest 3 Sound		Subtest 4 Rhythm	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A	1.82	.98	1.72	.82	1.39	.92	1.35	1.03
B	2.19	.95	1.33	.95	1.52	1.14	1.96	1.03

GRADE X								
Test Form	Subtest 1 Thought		Subtest 2 Suggestion		Subtest 3 Sound		Subtest 4 Rhythm	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A	1.55	.95	1.57	.86	1.33	.79	1.02	.89
B	1.98	.96	.94	.87	1.18	.98	1.71	.99

GRADE XII								
Test Form	Subtest 1 Thought		Subtest 2 Suggestion		Subtest 3 Sound		Subtest 4 Rhythm	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A	2.11	.93	1.89	.83	1.45	1.03	1.73	1.05
B	2.41	.90	1.71	.86	1.86	1.18	2.20	1.01

between Forms A and B than is revealed by the comparison of total test means presented in Table IV. These differences in subtest means provide a further indication that the test forms, though roughly comparable, are lacking in equivalence.

A comparison by inspection of Form A subtest means with the corresponding Form B subtest means produces a somewhat irregular pattern from which it is difficult to make clear-cut inferences regarding performance by each of the grade groups on the four subtests. This irregularity is not unexpected in view of the complexity of the skills measured by each subtest and in view of the fact that each subtest, consisting as it does of only four items, represents a very small sample from a total population of possible items. Thus, the Grade X mean for the Suggestion Subtest of Form B, for example, is inexplicably low in relation to the other Form B subtest means for Grade X.

An inspection of the data presented in Table XI suggests that there is a higher degree of similarity between the Form A and B means of three of the four subtests for Grade XII than for Grade X. This is perhaps due to the likelihood that the test items on the whole are too difficult, and therefore of questionable reliability, for Grade X students.

More central to the purposes of this study, however, is the question of the relative difficulty of each of the four subtests as it reflects the degree of skill possessed

by high school students in the area measured by that subtest. A true comparison of these difficulty levels, it must be pointed out, is not possible from the present data because mean scores for the various subtests reflect the widely differing difficulty levels of the small number of items comprising each subtest as much as they do skills of the students. Table XII indicates that each subtest contains both easy and difficult items. It is clear that the mean score for any one subtest could easily be dramatically raised or lowered through a change in the selection of constituent items.

TABLE XII

Item Difficulty Index Range Within Each
Subtest of Forms A and B

Subtest	Form A		Form B	
	Grade X	Grade XII	Grade X	Grade XII
1	.25-.63	.26-.72	.25-.65	.49-.75
2	.16-.49	.28-.48	.25-.39	.44-.51
3	.06-.74	.09-.80	.08-.45	.28-.55
4	.16-.33	.30-.58	.08-.63	.17-.70

Nevertheless, because the items comprising the final test forms were selected from over four times as many items on the basis of their discriminating power rather than their difficulty level, certain inferences concerning the relative abilities of students in the areas measured by the subtests

appear justified. A clearer if somewhat oversimplified picture of the relative difficulties of the four subtests is provided by the combined means of the two Test Forms presented in Table XI. It is granted that this procedure is statistically unsound because the two forms were administered to different groups. It does however permit a ranking of the four subtests (shown in Table XIII) from which certain generalizations can be made:

1. Students generally found Subtest One (Thought) easier than the other Subtests.
2. Students generally found Subtest Four (Rhythm) easier than Subtests Two and Three (Suggestion and Sound).
3. Students in the two grades responded more consistently to Form B than to Form A, and thus Form B would appear more generally reliable as a measuring instrument.

The first two of these generalizations corroborate the results of the pilot and pre-pilot studies, in which students generally found Subtest Two (Suggestion) and Subtest Three (Sound) items more difficult than items in Subtest One (Thought) and Subtest Four (Rhythm).

With the forementioned reservations in mind, certain other interpretations appear justified from the composite data presented in Table XI, which gives the means and standard deviations of the four subtest scores on each of the two test forms.

TABLE XIII

Rank Orders of Mean Subtest Scores
by Test Forms and Grade Groups

Source of Mean Scores	Subtest 1 Thought	Subtest 2 Suggestion	Subtest 3 Sound	Subtest 4 Rhythm
Form A Total	1	2	3	4
Form A Grade X	2	1	3	4
Form A Grade XII	1	2	4	3
Form B Total	1	4	3	2
Form B Grade X	1	4	3	2
Form B Grade XII	1	4	3	2
Combined Grade X	1	3.5	3.5	2
Combined Grade XII	1	4	4	2
Combined Total	1	4	4	2

1. As is shown in Table XIV, which is based on the data presented in Table XI, only for Subtest One (Thought) are the increases of Grade XII scores over Grade X scores similar for the two Test Forms. For the other subtests these increases differ substantially for the two Test Forms. The Subtest Four (Rhythm) increases appear somewhat more consistent than those for Subtest Two (Suggestion) and for Subtest Three (Sound), where increases as low as 9 per cent and as high as 82 per cent appear.

TABLE XIV

Per cent Increases of Mean Scores of Grade XII Over Grade X for the Four Subtests of Forms A and B

Increases Grade XII over Grade X = $\frac{\text{Grade XII} - \text{Grade X}}{\text{Grade X}} \times 100$				
	Subtest 1 Thought	Subtest 2 Suggestion	Subtest 3 Sound	Subtest 4 Rhythm
A	36	20	9	70
B	22	82	58	28
Av.	29	56	34	49

2. The data presented in Table XIV indicate that the percentage increases of Grade XII scores over Grade X scores are generally greater for Subtest Two (Suggestion) and Subtest Four (Rhythm) than for Subtest One (Thought) and Subtest Three (Sound). This matter bears further study, and is perhaps accounted for by the fact that Subtest Two

(Suggestion) and Subtest Four (Rhythm) relate to special technical skills more peculiar to poetry and hence more susceptible to short-term improvement than the skills measured by Subtest One (Thought) and Subtest Three (Sound). The components measured by Subtest One (the "Thought" or overall meaning of a poem) and by Subtest Three (the effects of the sounds of individual words) would seem more closely related to everyday linguistic skills than the more exclusively poetic components measured by Subtest Two (Suggestion) and Subtest Four (Rhythm).

3. The inferences made in the previous section must be viewed with caution since, as shown in Table XI, the means for Form B, Subtests Two (Suggestion) and Three (Sound), and Form A, Subtest Four (Rhythm), are extremely close to the chance level. As noted earlier, the Poetry Test, particularly in Subtests Two (Suggestion), Three (Sound), and Four (Rhythm), appears too difficult for most Grade X students to produce results having any appreciable degree of validity.

It appears, then, that the Poetry Test was successful only to a limited degree in revealing the relative competence of high school students in the areas measured by the subtests. The Test appears more suitable for Grade XII than for Grade X students and, in general, Form B appears more reliable than Form A.

That it is easier to find consistency in the results of the "Thought" Subtest than in the results of the

"Suggestion", "Sound", and "Rhythm" Subtests can perhaps be attributed to the possibility that the "Thought" Subtest measures a more general and hence stable skill more akin to general reading comprehension than the skills measured by the other subtests, which deal with uses of language more peculiar to poetic communication.

The irregularity of the pattern of results produced by the "Suggestion" and "Sound" Subtests and to a lesser extent the "Rhythm" Subtest is not unexpected in the light of the comments regarding complexity of poetic language reviewed in Chapter II and the difficulty in categorization described in Chapter III.

Item data for the four subtests. The section of Table XV which gives the mean item difficulty indices for each of the four subtests presents substantially the same data as is given in Table XI. Because each subtest consists of four items, multiplying the item difficulty indices by four produces roughly the mean subtest scores found in Table XI. These data have been discussed in the previous section and hence require no further comment at this point.

An examination of the second section of Table XV, which presents the mean biserial correlation coefficients for the four subtests in each Test Form, reveals that the overall discriminating power of each of the four subtests is about the same as that of the total test. Consistent

TABLE XV
Item Data Means for the Four Subtests for Forms A and B

	Thought		Suggestion		Sound		Rhythm	
	A	B	Mean	A	B	Mean	A	B
<u>1. Mean Difficulty Indices</u>								
Total Group	.45	.54	.50	.34	.38	.36	.41	.32
Grade X	.39	.49	.44	.31	.30	.31	.38	.24
Grade XII	.51	.61	.56	.38	.49	.44	.46	.42
<u>2. Mean Biserial Correlation Coefficients</u>								
Total Group	.40	.46	.43	.31	.60	.45	.49	.45
Grade X	.49	.36	.42	.25	.49	.37	.43	.28
Grade XII	.35	.41	.43	.37	.62	.49	.48	.44
<u>3. Mean Item Reliability Indices</u>								
Total Group	.19	.22	.20	.14	.29	.21	.21	.23
Grade X	.23	.17	.20	.10	.22	.16	.19	.10
Grade XII	.16	.24	.20	.18	.31	.24	.20	.21

with the data for the total Test Forms presented in Table II, there is a markedly greater increase in biserial correlation coefficients of Grade XII over Grade X for Form B than for Form A on all four subtests. These data appear to supply evidence that Form B is superior to Form A not only in overall discriminating power but also in the discriminating power of each of the four subtests.

The second section of Table XV further indicates that, with the exception of Subtest One (Thought), the subtest results confirm the generalization proposed earlier that the test items as a whole have more discriminating power for Grade XII than for Grade X students, and that the Test appears more reliable for Grade XII than for Grade X.

The third section of Table XV presents the mean item reliability indices for the four subtests of each Test Form. Unless the item difficulty index for an item deviates widely from the 0.5 level, the item reliability is closely related to the biserial correlation coefficient. Thus the data presented in this section confirm the similarity of the mean item discriminating power of the four subtests shown in section two of the table. It would appear that with the possible exception of "Suggestion" the Poetry Test as a whole was fairly successful in terms of the similarity of the mean item discriminating power of the four subtests in each of the two Forms.

Table VII (page 101) presents item data obtained by determining the correlations between performance on

individual items and subtest performance. As was noted earlier, the value of this procedure is curtailed by the small number of items in each subtest, making the biserial correlation coefficients and hence these item reliability indices spuriously high. This procedure, however, does provide data which corroborates that presented in Table XV. Again, with the possible exception of "Suggestion", a comparison of the figures both across the subtests and across the two test forms reveals a high degree of similarity. It is therefore possible to produce items of similar discriminating ability for each of the four subtests or categories. In this respect the present testing procedure appears practicable.

Intercorrelations among subtest scores. Intercorrelations among the four subtest scores were computed in an attempt to discover the extent to which the skills measured by the subtests are related to one another and possible patterns of relationships between the skills measured by the subtests.

Table XVI reveals that over-all, the intercorrelations among subtest scores are positive but low, suggesting that the skills measured by the subtests, while interrelated to some extent, are highly discrete. As shown in the table, for both the total group and the Grade XII group a positive correlation significantly different from zero occurs in seven out of twelve cases, while for the Grade X group no significant correlations occur. The table also shows that significant positive correlations occur five times for Form B and only

TABLE XVI
Intercorrelations Among Subtest
Scores for Forms A and B

	Subtest 1 Thought		Subtest 2 Suggestion		Subtest 3 Sound	
	A	B	A	B	A	B
TOTAL GROUP						
Subtest 1						
Subtest 2	.27**	.30**				
Subtest 3	.08	.32**	.10	.29**		
Subtest 4	.10	.24**	.26**	.12	.003	.32**
GRADE X						
Subtest 1						
Subtest 2	.19	.07				
Subtest 3	.14	.05	-.12	-.13		
Subtest 4	.11	.04	-.17	-.09	-.24	-.24
GRADE XII						
Subtest 1						
Subtest 2	.28*	.42**				
Subtest 3	-.01	.48**	.27*	.47**		
Subtest 4	-.08	.36**	.25	.14	.14	.28*

*p < 0.05

**p < 0.01

twice for Form A.

By converting coefficients to standard scores, it is possible to average two or more correlation coefficients.¹ As estimated by using this procedure, the correlations between all pairs of subtests of the Poetry Test Forms average less than 0.20 while the intercorrelations between scores on the two subtests of the Nelson-Denny Standardized Reading Test average over 0.60. Thus there appears to be a substantially lower degree of relationship between the component skills measured by the present Poetry Test than between skill in Vocabulary and Reading Comprehension as measured by the Nelson-Denny Test. This suggests that possibly the complex of skills comprising the construct "Skill in Responding to Poetic Language" which the experimental test seeks to measure is more diversified and complicated than what is usually thought of as General Reading Ability. These findings appear to lend support to the views expressed in previous chapters regarding the relative complexity of poetic language.

Further examination of Table XVI reveals wide differences between the various inter-subtest correlations for the two Forms. This reflects the differences in the correlations between the subtest scores and total scores presented in Table VIII (page 102) and substantiates the difficulties in categorization discussed in Chapter III. The present categorization scheme appears to be somewhat inadequate for coping with the complexities of poetic language and would seem to require further refinement and modification.

¹Quinn McNemar, Psychological Statistics (New York: Wiley, 1962, p. 140.

Further, the irregularity of the pattern of intercorrelations seen in Table XVI makes it difficult to formulate clear generalizations regarding the relationships between the various skills measured by the four subtests. Nevertheless, certain interpretations seem warranted on a tentative basis:

1. As Table XVII, which is a supplement to Table XVI, clearly shows, mean intercorrelations, as estimated by the z-transformation procedure, are much higher for the Grade XII group than for the Grade X group. More specifically, many of the intercorrelations for the Grade X group are either very low or negative and their average is close to zero. This is perhaps best accountable by the fact that the Grade X means are close to the chance level. This result, in addition to indicating the limited suitability of the test as a whole for Grade X students, would seem to cast considerable doubt on the significance of the statistical material and inferences based on Grade X data. For this reason the subsequent generalizations are based mainly on the Grade XII intercorrelations.

2. Table XVII also reveals that for the total sample and for the Grade XII group the estimated average inter-subtest correlation is substantially higher for Form B than for Form A. This appears to reflect the higher mean subtest scores presented in Table XI (page 106) and probably the greater reliability of Form B indicated by the data in Table I (page 93).

TABLE XVII
Mean¹ Intercorrelations among the Four Subtests
by Grade Groups and Test Forms

	A	B	Form Means
Total	.15	.26**	.21
Grade X	.04	.03	.04
Grade XII	.15	.42**	.29*
Grade Means	.09	.29*	.17

*p < 0.05

**p < 0.01

3. From the data presented in Table XVI for the Grade XII group, the highest intercorrelations are between Subtests One ("Thought") and Two ("Suggestion") and between Subtests Two and Three ("Sound"), while the lowest are between Subtests One and Four ("Rhythm"). The considerable difference between the correlations found in the Test Forms agrees with what might be expected in view of the nature of the skills measured by the subtests.

It seems reasonable, for example, to expect that the skill measured by Subtest One, which has to do with the "plain sense" meaning of a poem or part of a poem, will bear a closer relationship to skill measured by Subtest Two, which has to do with the "poetic", or suggestive, meaning of a word or expression as it relates to its context than

¹As estimated by means of the z-transformation technique (see McNemar, loc. cit.).

it will to the skill measured by Subtest Four, which has to do with sensitivity to poetic rhythms.

It also seems reasonable to expect that the skill measured by Subtest Three, sensitivity to the non-rhythmical sounds of individual words and expressions, will bear a closer relationship to the skill measured by Subtest Two ("Suggestion"), because the suggestive power of a word is reinforced by its sound, than it will to the skill measured by Subtest One ("Thought"). All of the foregoing expectations are borne out to some degree by the Grade XII data.

There is, however, one possible expectation which is not supported by the data presented in Table XVI. This is the expectation that there would be a relatively high correlation between Subtest Three ("Sound") and Subtest Four ("Rhythm") since each measures sensitivity to a kind of word-sound. According to the findings of this study, sensitivity to the sounds made by words appears to be a complex of fairly discrete skills rather than a single skill. This relationship bears further investigation.

In summary, the patterns of relationships between the skills measured by the four subtests are only dimly revealed by the present study. This is an exploratory type of investigation which not unexpectedly, in view of the complexity of poetic language, failed to produce an instrument sharply enough defined to distinguish clearly among the various skills the subtests attempt to measure. The study does, however, provide some direction for further

investigation. This matter will be discussed further in Chapter VI.

Comparison of subtest scores with reading scores.

Table XVIII presents the intercorrelations between each of the subtest scores of the two Poetry Test forms and the three Reading Test scores. Again differences between data produced by the two Forms and the irregular pattern, particularly in the Grade X data, make it difficult to form clear interpretations.

The data presented earlier (Table IX, page 104) showing low-positive correlations between totals of both Poetry Test forms and Reading Test totals suggest that the two tests measure related but reasonably discrete abilities. Because a central concern of this study is to examine the difference between response to discursive language and response to poetic language, or in other words, between skill in "reading comprehension" as it is normally measured, and skill in reading a poem, the intercorrelations among the four Poetry Subtests and three Reading Subtests are examined in some detail.

Of the total of 72 intercorrelations presented in Table XVIII, 39, or slightly more than half, are significantly different from zero. Of these 39, 20 result from Form A and 19 from Form B, suggesting that overall, the two test forms bear a similar relationship to the Reading Test. Also similar are the data from the two Grade Groups:

TABLE XVIII

Intercorrelations of Subtest Scores
and Reading Test Scores

Reading Score	Subtest 1 Thought		Subtest 2 Suggestion		Subtest 3 Sound		Subtest 4 Rhythm	
	A	B	A	B	A	B	A	B
<u>TOTAL GROUP</u>								
Vocabulary	.15	.42**	.30**	.15	.10	.24**	.48**	.44**
Comprehension	.30**	.07	.36**	.15	.07	.25**	.38**	.32**
Total	.24**	.29**	.35**	.17	.09	.28**	.47**	.43**
<u>GRADE X</u>								
Vocabulary	.21	.30*	.34**	.11	.12	.29*	.41**	.23*
Comprehension	.22	.23	.35**	.01	.12	.18	.36**	.02
Total	.23*	.01	.37**	.07	.13	.28*	.41*	.14
<u>GRADE XII</u>								
Vocabulary	.14	.45**	.15	.03	.01	.08	.29*	.54**
Comprehension	.27*	.27*	.28*	.12	.07	.21	.11	.54**
Total	.03	.41**	.23*	.07	.03	.15	.25*	.60**

*p < 0.05

**p < 0.01

for both Grade X and Grade XII 11 out of a total of 24 correlations are significantly different from zero. Thus the overall relationship between Reading Test performance and Poetry test performance appears to be approximately the same for Grade X and Grade XII students.

An examination of the intercorrelations for the individual Poetry Test subtests, however, reveals certain important disparities. As Table XIX which is derived from the Table XVIII data clearly shows, for the total sample, performance on the "Rhythm" Subtest correlates the most highly with Reading Test performance, while performance on the "Sound" Subtest correlates the least highly with Reading Test performance.

TABLE XIX

Significant Intercorrelations between Poetry
Subtest Scores and Reading Test Scores

Poetry Test	Numbers of Significant Correlations for Poetry Subtests (out of 9)			
	Subtest 1 Thought	Subtest 2 Suggestion	Subtest 3 Sound	Subtest 4 Rhythm
Form A	4	8	0	8
Form B	6	1	5	7
Total	10	9	5	15

For the Grade X group, very low correlations appear for Form B scores, particularly those for Comprehension and those for Subtest Two ("Suggestion"). For Grade XII the intercorrelations for Form B are somewhat higher than those for Form A, except for the "Suggestion" Subtest. The "Thought" and "Rhythm" Subtests correlate more highly with Reading Test scores than do scores on the "Suggestion" and "Sound" Subtests.

Certain inferences and explanations consistent with these data provide further evidence that Forms A and B are not equivalent. It has been tentatively concluded that, because Grade X scores are often very close to chance level, Grade XII results are more reliable. It has also been tentatively concluded from the item data that Form B is more reliable as a source of inferences than Form A. An examination of the Grade XII scores, particularly those for Form B, reveals that performance on Subtest Four ("Rhythm") correlates more highly with Reading Test scores than does performance on the other subtests. This correlation is followed fairly closely by Subtest One ("Thought"), with much lower correlations for Subtests Two ("Suggestion") and Three ("Sound"). The correlation coefficients for Subtests Four, One, Three, and Two, are 0.60, 0.47, 0.15, and 0.07 respectively. The last two are not significantly different from zero.

An examination of Table XIX and the figures on Table XVII on which it is based reveals a pronounced pattern

in performance on the "Suggestion" and "Sound" subtests of Forms A and B for which it is difficult to account. According to this pattern for both Grade X and Grade XII students there is a significant relationship between scores on Subtest Two ("Suggestion") of Form A and Reading Test scores and virtually no relationship between scores on Subtest Two of Form B and Reading Test scores. For Subtest Three ("Sound") this pattern is exactly reversed: Form B scores are significantly related and Form A scores unrelated to Reading Test scores.

A close investigation of the relevant items and an attempt to relate this pattern to the factor loadings discussed in a later section failed to account for this pattern. It can only be attributed to the possibility that because of the very small sample of items in each subtest and the complexity of the skills being measured, the four items in Subtest Two Form A and the four items in Subtest Three Form B unintentionally measure, to a substantial degree, skills common to all reading. Conversely, it appears that the four items in Subtest Two Form B and the four items in Subtest Three Form A unwittingly measure skills more idiosyncratic either to the test itself or to poetic language in general.

Another pattern is revealed by the figures presented in Table XX, which is also derived from Table XVIII. There is some indication here that Form A items generally operate more at the poem's total meaning level, while Form B items

operate more at the single-word-meaning level. Because the intention of the test was to stress the latter and because most of the data presented to this point suggest that B is the more successful form in terms of this intention, the pattern suggested by Table XIX is not unexpected.

TABLE XX

Significant Intercorrelations between Reading
Subtest Scores and Poetry Test Scores

Poetry Test Form	Number of Significant Correlations for Reading Subtests (out of 12)	
	Vocabulary	Comprehension
A	5	7
B	8	4

A further anomaly revealed by Table XVIII is presented in simplified form in Table XXI. For both Subtest One ("Thought") and Subtest Four ("Rhythm") there is a noticeable Grade X-Grade XII increase in Poetry Test-Reading Test intercorrelations for Form B. Conversely, there is a corresponding decrease for Form A. In the case of Subtest One ("Thought"), this irregularity is perhaps due to the fact that three of the four Form A items were notably unsuccessful in requiring students to make inferences about the "thought" of a poem. With regard to Subtest Four ("Rhythm"), an explanation appears to be that the Form B items tended to stress the metrical pattern of the whole

poem while the Form A items tended to involve metrical nuances within individual lines. Thus the Form B "Rhythm" items probably measured a skill more analytical, and hence more closely related to "reading comprehension", than the corresponding Form A items.

TABLE XXI

Intercorrelations between Reading Test
Scores and Scores on Subtests One
and Four by Grade Level

Grade Level	Subtest 1 Thought		Subtest 4 Rhythm	
	Form A	Form B	Form A	Form B
X	.23	.01	.41	.14
XII	.03	.41	.25	.60

Of the skills measured by the four subtests, that measured by Subtest One, which attempted to deal with the "thought" or "overall meaning" of a poem, might have been expected to correlate the most highly with the skills measured by a standardized reading test. Contrary to expectation, this skill correlated less highly with Reading Test scores than did the skill measured by Subtest Four for the Grade 12 group, and correlated the least highly of all with subtest scores with Reading Test scores for the Grade X group.

It would therefore appear that the kind of reading

skill measured by Subtest One ("Thought") is different from general reading skill as measured by standardized reading tests. This difference has a tendency to diminish with older students.

The generally low Grade XII correlations for Subtests Two ("Suggestion") and Three ("Sound") may be explained by suggesting that these subtests measure skills largely idiosyncratic to poetic language. It is more difficult, however, to explain the relatively high correlations between Subtest Four ("Rhythm") results and Reading Test results. It would seem that Subtest Four, like Subtests Two and Three, measures something more or less peculiar to poetic language (rhythm and metre). On the face of it, a low correlation would be expected here. Perhaps the relatively high correlations may result from a common skill, somewhat analytical in nature, which underlies both ability to discern a metrical pattern and to succeed in a standardized reading test. The factor analysis results presented later in this chapter as well as the interview results presented in chapter V, tend to bear out this possibility, which appears a fruitful avenue for further study.

A final, highly tentative conclusion arising from the data presented in Table XVIII has to do with a comparison of Grade X and Grade XII coefficients for the four subtests. This comparison reveals a substantial general gain between Grade X and Grade XII in performance on subtests One and Four, a small gain in performance on Subtest Two,

and no gain in performance on Subtest Three. A possible explanation is that, like the Reading Test, Subtests One, Two and Three measure skills more capable of fuller development during adolescence than is sensitivity to the flavour and sounds of words. Thus the highest correlations in Table XVIII are between the Vocabulary section of the Reading Test and Subtest Four ("Rhythm"), both of which measure skills capable of rapid development, and the lowest correlations, at least for the Grade XII group, are between Reading scores and Subtest Three ("Sound") scores. Apparently sensitivity to word sounds does not, like skills measured by Subtests One, Two, and Four, become more closely related to general reading skill, as measured by a standardized reading test, as students mature. Perhaps this points to a deficiency in standardized reading tests, and suggests their neglect of certain kinds of word meanings connected with nuance and sound. There seems little doubt that this matter requires further investigation.

Finally, it appears that the anomalies revealed in Tables XIX, XX, and XXI, in defying easy interpretation, provide further evidence that each Subtest and indeed each item of the Poetry Test does not measure a single simple skill but rather a complex of skills which is difficult to analyze on the basis of the present data.

Summary of Characteristics of Test Forms and Subtests

Before proceeding to report in some detail on

performance on individual Poetry Test items, it seems useful to present a brief review of the main findings reported thus far. These findings, which relate to the Test Forms as a whole and to the Four Subtests, are presented as follows:

1. In view of the high degree of complexity of the skills being measured, and in view of its purposes as an investigating instrument (see Guilford's comment on page 96) the Poetry Test was found to be satisfactorily reliable.

2. The two test forms were found to be comparable, but were not found to meet standards for statistical parallelism acknowledged by current authorities on testing. This lack of equivalence is particularly evident in comparing the difficulty levels and discriminating power between each of the four pairs of subtests.

3. Most of the data suggests that Form B is generally more reliable than Form A.

4. The Poetry Test generally appears more reliable for Grade XII students than for Grade X students. Also, in terms of difficulty, the Poetry Test is considerably more suitable for Grade XII than for Grade X students.

5. Subtest One ("Thought") results and Subtest Four ("Rhythm") results show a more consistent performance pattern than do Subtest Two ("Suggestion") and Subtest Three ("Sound") results.

6. Of the four subtests, Subtest One ("Thought") was generally the least difficult, and Subtests Two ("Suggestion") and Three ("Sound") were generally the most

difficult.

7. There was a greater performance increase between Grades X and XII with Subtests Two and Four than with Subtests One and Three.

8. The mean item discriminatory power is approximately the same for the four subtests and for the two test forms.

9. There is generally a low level of intercorrelation among scores on the four subtests (mean correlations < 0.20). The degree of intercorrelation is higher for the Grade XII students than for the Grade X students.

10. The correlations between performance on Subtests One ("Thought") and Two ("Suggestion") and between Subtests Two and Three ("Sound") are relatively high; the correlation between performance on Subtests Three and Four ("Rhythm") is relatively low.

11. There is a low positive correlation between Poetry Test performance and Reading Test performance. This correlation is approximately the same for Grade X and Grade XII.

12. For Form B, which evidence suggests is the more reliable of the two Poetry Test forms, the difference between Grade X and Grade XII performance appears to be greater for Poetry Test scores than for Reading Test scores.

13. Generally, performance on the "Rhythm" Subtest has the highest correlation and performance on the "Sound" Subtest the lowest correlation with Reading Test performance. There is also a relatively low degree of correlation between

performance on the "Suggestion" Subtest and performance on the Reading Test.

14. The correlations between performance in the four subtests and the Reading Test generally increased from Grade X to Grade XII. Of these, the smallest increase is for the "Sound" Subtest.

Characteristics of Individual Items

Complete item data for forms A and B. Tables XXII and XXIII present the difficulty indices, biserial correlation coefficients, and item reliability indices of all items on both Test Forms. Much of this material has been summarized in Tables II, VI, VII, IX, and XV and discussed in the two previous sections. However, summary data involving means of groups of items can fail to reveal specific characteristics of individual items.

Guilford points out that "it is more important to analyze aptitude tests than achievement tests," and that "in achievement tests it is sometimes more important to have the items approved by a subject matter expert than to know their correlation with the total score."¹ Because the present test is as much an achievement as an aptitude test, it would perhaps not seem reasonable to apply item analysis criteria too rigorously, particularly in cases where it conflicts with the judgment of the investigator and the judges. The present test, in attempting to measure skill

¹J. P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill, 1965), p. 446.

TABLE XXII

Comparison of Item Data for Grade X and XII
by Subtests for Form A

Item Number	Difficulty		Biserial Corr.		Item Rel. Index	
	Gr. X	Gr. XII	Gr. X	Gr. XII	Gr. X	Gr. XII
<u>SUBTEST 1: THOUGHT</u>						
2	.29	.48	.27	.32	.12	.16
3	.25	.72	.84	.05	.36	.02
10	.63	.59	.34	.42	.17	.21
11	.41	.26	.52	.59	.26	.26
Means	.39	.51	.49	.35	.23	.16
<u>SUBTEST 2: SUGGESTION</u>						
4	.49	.46	.26	.13	.13	.07
6	.29	.30	.21	.54	.09	.25
12	.16	.48	.38	.56	.14	.28
15	.30	.28	.06	.26	.03	.12
Means	.31	.38	.25	.37	.10	.18
<u>SUBTEST 3: SOUND</u>						
1	.74	.80	.56	.36	.25	.14
8	.45	.48	.62	.56	.31	.28
9	.29	.48	.30	.49	.14	.24
14	.06	.09	.25	.52	.06	.15
Means	.38	.46	.43	.48	.19	.20
<u>SUBTEST 4: RHYTHM</u>						
5	.33	.37	.37	.47	.18	.23
7	.29	.41	.30	.43	.14	.21
13	.16	.30	.46	.60	.17	.28
16	.28	.58	.32	.22	.14	.11
Means	.26	.42	.39	.43	.16	.21

TABLE XXIII

Comparison of Item Data for Grade X and XII
by Subtests for Form B

Item Number	Difficulty		Biserial Corr.		Item Rel. Index	
	Gr. X	Gr. XII	Gr. X	Gr. XII	Gr. X	Gr. XII
<u>SUBTEST 1: THOUGHT</u>						
1	.65	.66	.46	.52	.22	.25
2	.47	.53	.15	.47	.08	.23
3	.61	.75	.46	.52	.22	.23
13	.25	.49	.37	.54	.16	.27
Means	.49	.61	.36	.51	.17	.24
<u>SUBTEST 2: SUGGESTION</u>						
8	.25	.51	.50	.71	.21	.35
9	.27	.47	.57	.51	.25	.25
11	.27	.45	.48	.71	.21	.36
16	.39	.44	.42	.55	.21	.27
Means	.30	.47	.49	.62	.22	.31
<u>SUBTEST 3: SOUND</u>						
5	.16	.38	.40	.71	.15	.34
6	.45	.55	.12	.33	.06	.17
12	.08	.28	.32	.23	.09	.10
15	.26	.46	.28	.49	.12	.25
Means	.24	.42	.28	.44	.10	.21
<u>SUBTEST 4: RHYTHM</u>						
4	.63	.64	.47	.56	.23	.27
7	.51	.70	.39	.43	.20	.20
10	.49	.62	.51	.41	.25	.20
14	.08	.17	.32	.84	.09	.32
Means	.43	.53	.42	.56	.19	.25

in responding to poetic language, deals with a complex of skills or abilities rather than a single skill or ability. In its construction according to the categories reported in Chapter III, the test acknowledges this heterogeneity.

Thus, useful as item analysis data may be in helping the test maker select suitable items, it needs to be interpreted and applied with a certain measure of flexibility in the present study. Items which during the pilot studies showed a difficulty index of less than 0.30 or more than 0.70 were for the most part discarded. Similarly, items with a biserial correlation coefficient of less than 0.35 were generally discarded. As Table XXIV shows, many items as administered to the Grade X group and several items as administered to the Grade XII group failed to meet even these reasonably flexible criteria in the final test administration. Table XXIV provides further evidence that the Poetry Test as a whole is more suitable for Grade XII than for Grade X students. The two Test Forms are fairly similar in the distribution of items by difficulty level and discriminatory ability although Form B has somewhat greater discriminatory power at the Grade XII level than Form A.

In presenting the range of item data figures for the four subtests in each form, Tables XII (page 108) and XXV indicate the extent to which individual items deviated from the desired difficulty and discrimination levels. Table XII (page 108) shows items in all subtests except Subtest

TABLE XXIV

Range of Biserial Correlation Coefficients of
Items within Each Subtest of Forms A and B

Subtest	Grade X		Grade XII	
	A	B	A	B
1	.27-.84	.15-.46	.05-.59	.47-.54
2	.06-.38	.42-.57	.13-.56	.55-.71
3	.25-.56	.12-.40	.36-.56	.33-.71
4	.30-.46	.32-.51	.22-.60	.41-.84

TABLE XXV

Distribution of Items by Difficulty Level and
Discriminatory Power Number of Items out of 16

Grade and Test Form	Difficulty Index			Biserial Correl.		
	.40-.60	.30-.70	.20-.80	.50+	.35+	.20+
X A	3	5	13	4	7	15
X B	4	8	13	2	11	14
XII A	8	11	14	6	11	14
XII B	8	12	15	10	14	16

One ("Thought") as falling outside the desired difficulty range. Table XXIV shows all subtests excepting possibly Subtest Four ("Rhythm") as including at least one item with an extremely low biserial correlation coefficient.

The purpose of Table XXVI is to indicate the extent to which difficulty indices and biserial correlation coefficients changed for identical items between the pilot study and the final test administration. Four items from each test form were selected to illustrate the substantial changes in both difficulty level and discriminating power. Approximately half the items in each form showed substantial changes of this nature. Although these changes could perhaps have been minimized by using a larger sample for the pilot study (only two classes were used), the comparison presented in Table XXVI points out the difficulty of creating two equivalent test forms. This comparison also raises doubts as to whether the mode of questioning developed for the present experimental test lends itself to a sufficient degree of reliability to render such a test useful as an evaluative as opposed to a heuristic or research instrument. More research involving the refinement of the questioning technique and the administration of the items to larger and more diverse groups would seem to be necessary before more definite conclusions can be reached in this matter. The skills which these items attempt to measure appear to be highly complex, involving so many variables that even minute changes in the form of an item

TABLE XXVI

Comparison of Item Data from the Pilot Study
and from the Final Test Administration for
Selected Items from Forms A and B

Item No.	Item Difficulty		Biserial Corr.	
	Pilot	Final	Pilot	Final
<u>FORM A ITEMS</u>				
2	.81	.37	.37	.37
4	.53	.47	.46	.18
5	.56	.34	.65	.42
7	.66	.34	.39	.41
8	.78	.46	.77	.57
10	.75	.61	.57	.31
13	.20	.23	.30	.59
14	.25	.07	.45	.42

TABLE XXVI (Continued)

Item No.	Item Difficulty		Biserial Corr.	
	Pilot	Final	Pilot	Final
<u>FORM B ITEMS</u>				
1	.80	.66	.24	.48
5	.43	.27	.69	.69
7	.70	.60	.65	.44
8	.43	.38	.34	.70
9	.38	.36	.37	.56
10	.63	.35	.31	.45
11	.47	.36	.35	.67
14	.40	.13	.56	.73
15	.23	.35	.17	.46

can produce substantial changes in the item data.

Detailed examination of response patterns for selected items. It was felt that the reporting of the results would be incomplete without some type of detailed examination of data relating to individual items. This kind of examination promises to provide information about the ways different sorts of pupils respond to items and about the ways slight changes in the form of items result in substantial changes in response patterns--information not adequately provided by statistics relating to groups of items alone. From this type of detailed information two kinds of inferences can arise: those concerning further directions for the construction of this general type of item, and those concerning specific student response to certain specific poetic stimuli.

Because of considerations of time and space, only 10 of the 32 items comprising both test forms have been selected for detailed discussion. This selection was difficult and of necessity somewhat arbitrary because all of the items used on the final test and many used in the pilot tests are amenable to the same kind of examination. An important consideration in the selection of the following ten items was an effort to reflect the various aspects of poetic language the test attempted to measure such as tone, rhythm, and imagery. Consequently, the following examination is organized according to these aspects of poetic

meaning.

(1) Tone

The "tone" of a poem has been defined by Richards as the manner in which the poet's language reflects his attitude toward his reader.¹ A perhaps more satisfactory definition is given by Perrine: "Tone, in literature, may be defined as the writer's attitude toward his subject, his audience, and himself. It is the emotional colouring or the emotional meaning of the work."² Obviously, the student's responsiveness to this "emotional meaning" is essential to the right reading of a poem. As Perrine and others have noted, however, while it is relatively easy to interpret tone in spoken language, the correct determination of tone in literature, which is indicated by a complex combination of elements such as connotation, imagery, irony, rhythm, and syntax, is a far more difficult matter, and depends on a high degree of familiarity with all of these elements. Thus in the present test, although it was considered necessary to include "tone" items, it was difficult to place such items under any one of the four existing categories. The most appropriate seemed to be Category Two, "Suggestion".

Two of the Poetry Test items attempted to measure students' awareness of "tone". In each of these items

¹I. A. Richards, Practical Criticism (New York: Harcourt, Brace, 1929), pp. 175, 197.

²Laurence Perrine, Sound and Sense (New York: Harcourt, Brace, 1963), p. 135.

("Portrait VIII" and "The Rich Man") an attempt was made to vary the distractors mainly on the "tone" dimension, keeping other elements such as "plain sense meaning" and metre as constant as possible.

(a) "The Rich Man"

Table XXVII presents detailed item data for Item 8 from Form B of the final test. To facilitate interpretation the entire item is included in the table.

In this table and in the nine succeeding tables the main figures to note, aside from the difficulty indices and the biserial correlation coefficients, are those giving the per cent of the total number of respondents selecting each alternative (given by the figures to the right of "Total") and the test score means of students selecting each alternative. For example, of the 96 students answering Item B-8, 53 per cent selected alternative two, the correct answer (indicated by an asterisk), and 27 per cent selected alternative three. The mean Poetry Test score achieved by the 53 per cent selecting the correct answer was 8; for the 27 per cent selecting alternative three, it was 5.5. Table XXVII also presents a breakdown of the total responding group of 96 students into fifths. The number "5" under "Selecting Group" refers to the twenty per cent of the examinees obtaining the highest test scores; "4" refers to the next twenty per cent, and so on down to "1", which refers to the twenty per cent obtaining the lowest

TABLE XXVII

Detailed Item Data for Item B-8¹

The rich man has his motor car,
His country and his town estate.
He smokes a fifty-cent cigar
and jeers at Fate.

He frivols through the livelong day,
He knows not Poverty her pinch.
His lot seems light, his heart seems gay,
He has a cinch.

Yet thought my lamp burns low and dim,
Though I must slave for livelihood -
Think you that I would change with him?

1. I wish I could.
2. You bet I would!*
3. Oh, if I could!
4. I think I would!

* = correct response

Series 1 Difficulty .53 Biserial Corr. .65

Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	<u>07</u>	<u>53</u>	<u>27</u>	<u>10</u>
5	14	86	0	0
4	0	50	50	0
3	0	67	33	0
2	13	50	25	13
1	0	14	43	29
Total Means	7.5	8.0	5.5	4.7

¹F. P. Adams, "The Rich Man," in Sound and Sense, ed. Laurence Perrine (New York: Harcourt, Brace and World, 1956), p. 40.

TABLE XXVII (Continued)

Pilot	Difficulty .43	Biserial Corr. .34			
Selecting Group	% Choosing Alternatives				
	1	2*	3	4	
Total	<u>20</u>	<u>43</u>	<u>27</u>	<u>10</u>	
5	0	71	14	14	
4	33	33	33	0	
3	30	30	30	10	
2	33	67	0	0	
1	14	29	43	14	
Total Means	8.1	10.4	8.6	10.3	

Final-Total Group	Difficulty .38	Biserial Corr. .70			
Total	<u>18</u>	<u>38</u>	<u>26</u>	<u>18</u>	
5	06	88	0	06	
4	14	36	4	36	
3	06	63	25	06	
2	26	19	35	19	
1	29	0	47	24	
Total Means	5.8	8.9	5.2	6.1	

TABLE XXVII (Continued)

Final - Grade X Difficulty .38 Biserial Corr. .70

Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	<u>22</u>	<u>24</u>	<u>29</u>	<u>24</u>
5	25	38	0	38
4	0	63	25	13
3	36	14	36	38
2	0	25	38	21
1	36	0	36	27
Total Means	5.5	7.1	5.0	5.7

Final - Grade XII Difficulty .51 Biserial Corr. .70

Total	<u>15</u>	<u>51</u>	<u>23</u>	<u>11</u>
5	0	89	0	11
4	17	58	7	8
3	9	64	18	9
2	43	29	29	0
1	13	0	63	25
Total Means	6.9	9.8	5.5	7.2

test scores. In this way these figures provide information both as to the relative attractiveness of each of the alternatives to the responding group as a whole and as to the type of student, in terms of test performance, selecting each alternative.

With reference to Table XXVII and the nine following tables, it could also be noted that "Series 1" and "Series 2" refer to the "pre-pilot" studies and that "Series X" refers to an amended version of these original series, in which an attempt was made to improve faulty items. Each of the test administrations previous to the final one involved two classes, or approximately 60 students. In the final test administration, as reported by these tables the N is 95 for Form A and 96 for Form B.

As is shown in Table XXVII, this item was presented in identical form to four groups of students at three different times. The ironical, self-mocking tone of the poem used for the item comes to a climax in the last line, "You bet I would!" In constructing the item the intent was to provide distractors at odds with this tone in various ways. Thus, "I think I would," the fourth alternative, was intended to project a neutral, or matter-of-fact tone, while "I wish I could!" and "Oh, if I could" were intended to project an increasing amount of fervour incompatible with the tone of the poem as a whole. Thus a "tone continuum" for the item would have alternative 2, the correct answer (ironic) at one end, alternative 4 (matter-of-fact) next,

alternative 1 (slightly fervid) next, and alternative 3 (fervid) at the opposite end of the continuum. As shown in Table XXVII, there is a marked preference for alternatives 2 and 3, those with more tone colour, over the more neutral alternatives 1 and 4, in four of the five responding groups. Alternative 2, the original (correct answer) was chosen by roughly half the respondents, with the exception of the Grade X group. The item was considered satisfactory in terms of its difficulty level. Alternative 3, furthest on the above tone continuum from the original, was the second most popular choice (Grade X excepted) and alternatives 1 and 4, the most neutral, were clearly the least popular.

The overall test means of students selecting the various alternatives presented in Table XXVII confirm the supposition that the number of persons selecting an alternative is probably, though not necessarily, related to the kind of persons making the selection. In this item, alternative 3, the second most popular choice, appears to have been chosen primarily by persons least skilled in responding to a poem. It is noteworthy that this alternative is furthest on the "tone continuum" from the correct answer. Similarly, there is a strong tendency for alternative 4, one of the least popular responses, but closest to the correct answer on the above continuum, to have been chosen by students with higher test scores than those selecting the other two distractors. In this way the item data correspond to the conceptual framework used in the

construction of the item and indicate the potential usefulness of information to be provided by this kind of procedure in diagnostic and research situations.

Another kind of information provided by Table XXVII is obtained by comparing the Grade X and Grade XII results for the final test administration. These figures indicate a marked difference between the sensitivity of Grade XII students and Grade X students to the tone of a poem. The extent to which this difference is due to the study of poetry in school, as opposed to other influences such as maturation and general experience in verbal communication, appears to be a matter worth investigating.

In summary, it appears that the item in question is reasonably successful in measuring what it set out to measure, and that scrutiny of detailed item data can provide information, particularly relating to "wrong" answers, which a summary examination of correct responses cannot provide. This detailed information promises to be of value in (a) refining this kind of item construction, (b) diagnosing specific teaching and learning difficulties, and (c) contributing to existing knowledge of response to poetry by high school students. For example, it seems clear that most members of the Grade X group responding to this item are unable to respond adequately to "tone" in this kind of poem, and perhaps in poems in general. More investigation would be needed to confirm this.

(b) "Portrait"

Table XXVIII presents item data for Item 12, Form A. As with the previous item, this item was presented to four groups of students at three different times. The poem on which it is based is more difficult than the one just discussed and the considerations affecting its tone are more subtle. Nevertheless an attempt was made to construct distractors not of varying attractiveness but of varying discriminating power. Thus an attempt was made to make alternative 4 (in its first-person appeal) most similar in tone to the original, and to make alternative 2 (in its theatrical extravagance) least similar.

The relative subtlety of what is being measured is reflected in the difficulty level figures. Only the Grade XII group, as a whole, seemed able to cope with the item at all adequately. Alternatives 1 and 4 were considerably more attractive than alternative 3, the "correct" one, for the Grade X group. Alternative 4, constructed as the second best choice, was selected more often than the two other distractors by the Grade XII group. Alternative 2, constructed as the worst choice, was selected least often by all groups.

Alternative 2 appeared to attract the least able poetry readers, the "correct" answer (alternative 3) the most able readers, and alternative 4 the second most able readers. These results suggest that the attempt to vary discriminating power of the distractors was generally

TABLE XXVIII (Continued)

Pilot	Difficulty .22		Biserial Corr. .33	
Selecting Group	% Choosing Alternatives			
	1	2	3*	4
Total	25	13	22	4
5	29	14	29	29
4	0	25	25	50
3	18	18	27	36
2	0	0	33	67
1	57	0	0	43
Total Means	9.4	11.3	11.4	10.0

Final - Total Group	Difficulty .31		Biserial Corr. .59	
Total	30	10	31	28
5	13	0	60	27
4	25	07	43	25
3	0	20	80	0
2	33	13	17	38
1	50	17	04	29
Total Means	5.2	4.7	7.7	6.0

TABLE XXVIII (Continued)

Final - Grade X Difficulty .16 Biserial Corr. .39

Selecting Group	% Choosing Alternatives			
	1	2	3*	4
Total	43	14	16	27
5	50	0	33	17
4	31	15	31	23
3	42	17	08	33
2	42	08	08	42
1	67	33	0	0
Total Means	5.2	4.6	6.5	5.4

Final - Grade XII Difficulty .48 Biserial Corr. .56

Total	15	07	48	30
5	0	0	75	25
4	15	0	62	23
3	0	13	63	25
2	25	08	25	42
1	40	20	0	40
Total Means	5.6	5.0	8.1	6.6

successful.

In comparing the Grade X with the Grade XII data it is seen that with the Grade XII group results, unlike the corresponding results in the previously discussed item, the order of attractiveness of the distractors parallels the order of discriminatory power, while with the Grade X group there appears to be little correspondence. The Grade XII data for this item represent the type of results which were hoped for in the construction of the Poetry Test items.

The Grade X data, on the other hand, indicate that although the item appears to have good discriminating power for Grade X students, it is generally too difficult for this level and, more important, that a sizeable difference may exist between the average Grade X student and the average Grade XII student in sensitivity to a poem's tone. This is consistent with the conclusion reached for the previous item, which also attempted to measure response to tone.

(2) Rhythm and Metrical Pattern

Of the 32 items in the Poetry Test, eight attempted to measure students' understanding of and sensitivity to poetic rhythm and to metrical patterns. It was originally expected that high school students would generally find the "metrical" items which comprise Subtest Four easier than other kinds of items, but this expectation was not borne out by the item data. Two of these metrical items, selected more or less at random, are examined in detail.

(a) "The Pulley"

Table XXIX presents detailed item data for Item 7, Form B, based on George Herbert's poem "The Pulley". The intent of this item was to measure the ability to detect a fairly obvious metrical pattern. Since each stanza ends with a trimeter line, by deleting the final line of one stanza and supplying as distractors lines with different numbers of feet or with feet which are irregular, it was hoped to measure students' awareness of metrical pattern. Thus, with other poetic elements kept as constant as possible, alternative 2 was made an irregular tetrameter line, alternative 3 a regular tetrameter line, and alternative 4 a pentameter line. It was expected that selection of alternative 3 would indicate a higher degree of awareness of metrical pattern than selection of alternatives 2 or 4.

Table XXIX indicates that this was a relatively easy item for all groups, with, as would be expected, Grade XII students finding it easier than Grade X students. Alternative 3 was the second choice for all groups, indicating that response to the item reflected the conceptual framework underlying its construction. An interesting point is that all groups showed a preference for the non-metrical tetrameter line over the metrical pentameter line, indicating that line length had more to do with the selection than metrical flow.

An examination of the test means suggests that, relatively speaking, the Grade XII students appear to be

TABLE XXIX

Detailed Item Data for Item B-7¹
(last 2 stanzas only)

For if I should (said He)
Bestow this jewel also on my creature,
He would adore my gifts instead of me,
And rest in Nature, not the God of Nature:
So both should losers be.

Yet let him keep the rest;
But keep them with repining restlessness;
Let him be rich and weary, that at last,
If goodness lead him not, yet weariness

-
1. May toss him to my breast*
 2. May yet toss him to my breast
 3. May toss him finally to my breast
 4. May finally toss him to my patient breast

Series 2 Difficulty .56 Biserial Corr. .42

Selecting Group	% Choosing Alternatives			
	1*	2	3	4
Total	54	13	19	10
5	75	8	8	0
4	43	29	14	14
3	50	10	40	0
2	73	9	0	18
1	25	17	33	7
Total Means	10.6	9.0	8.4	7.6

¹George Herbert, "The Pulley," in The English Works of George Herbert, ed. G. H. Palmer (Boston: Houghton-Mifflin, 1915), p. 149.

TABLE XXIX (Continued)

Pilot	Difficulty .70	Biserial Corr. .65			
Selecting Group	% Choosing Alternatives				
	1*	2	3	4	
Total	70	10	17	3	
5	100	0	0	0	
4	100	0	0	0	
3	70	20	10	0	
2	33	0	33	33	
1	43	14	43	0	
Total Means	10.4	8.7	7.0	8.0	

Final - Group Total	Difficulty .60	Biserial Corr. .44		
Total	60	13	17	11
5	88	6	0	6
4	86	0	14	0
3	38	25	13	25
2	52	13	29	6
1	47	18	18	18
Total Means	7.7	5.7	5.5	5.8

TABLE XXIX (Continued)

Final - Grade X Difficulty .51 Biserial Corr. .39

Selecting Group	% Choosing Alternatives			
	1*	2	3	4
Total	51	16	18	14
5	75	13	13	0
4	50	25	0	25
3	64	7	21	7
2	25	25	38	13
1	36	18	18	27
Total Means	6.4	5.6	5.1	4.7

Final - Grade XII Difficulty .70 Biserial Corr. .43

Total	70	9	15	6
5	89	0	0	11
4	92	0	8	0
3	45	18	18	18
2	43	14	43	0
1	75	13	13	0
Total Means	8.7	5.8	6.0	8.7

more influenced by metrical regularity while the Grade X students appear to be more influenced by line length. This matter clearly requires further investigation. Where the difficulty level is as high as 0.70, as it is in some of these data, examining second and third choices is not very productive because of the likelihood that persons failing to select the correct answer are operating at the chance level. However, it appears from these data that the item is reasonably successful in discriminating between students who are able to discern a fairly obvious metrical pattern and those who are not.

(b) "Tintern Abbey"

Table XXX presents detailed data for Item 7, Form A, based on an excerpt from Wordsworth's famous poem. As distractors for the original standard pentameter line (alternative 2), alternative 3 is an iambic pentameter line with a feminine ending, alternative 1 is a hexameter line, and alternative 4 is an irregular tetrameter line. It was expected that selection of alternative 1, even though not the correct answer, would indicate a higher level of discrimination than selection of alternative 3, and that selection of alternative 4, the least metrical of the four, would indicate the lowest level of metrical discrimination.

As indicated in Table XXX, these expectations are not borne out by item results, which are contradictory and indicate no clear response patterns. Again, this is

TABLE XXX
Detailed Item Data for Item A-7¹

And I have felt
A presence that disturbs me with the joy
Of elevated thoughts; a sense sublime
Of something far more deeply interfused,
Whose dwelling is the _____,
And the round ocean and the living air
And the blue sky and in the mind of man,
A motion and a spirit that impels
All thinking things, all objects of all thought,
And rolls through all things.

1. fading light of setting suns
2. light of setting suns*
3. fading light of sunsets
4. light of the sunset

Series 1 Difficulty .50 Biserial Corr. .33

Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	10	50	27	13
5	14	86	6	0
4	0	50	0	50
3	0	17	50	33
2	13	38	38	13
1	14	57	29	0
Total Means	6.3	7.5	5.8	7.0

¹Wm. Wordsworth, "Lines Composed a Few Miles Above Tintern Abbey," in Poetic Experience, ed. B. C. Diltz (Toronto: McLelland and Stewart, 1955), p. 278.

TABLE XXX (Continued)

Pilot	Difficulty .65		Biserial Corr. .39	
Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	25	66	6	3
5	14	86	0	0
4	25	75	0	0
3	27	55	9	9
2	0	100	0	0
1	43	43	14	0
Total Means	9.1	10.9	9.0	11.0

Final - Total Group	Difficulty .33		Biserial Corr. .41	
Total	29	34	14	22
5	33	60	0	7
4	18	39	14	25
3	40	60	0	0
2	33	21	25	21
1	33	21	13	33
Total Means	5.9	7.2	5.4	5.4

TABLE XXX (Continued)

Final - Grade X	Difficulty .29		Biserial Corr. .30	
Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	31	29	16	22
5	17	50	17	17
4	23	31	23	15
3	33	25	17	25
2	42	25	0	25
1	33	17	17	33
Total Means	4.8	6.1	5.8	4.8

Final - Grade XII	Difficulty .41		Biserial Corr. .43	
Total	26	41	11	22
5	25	75	6	0
4	38	38	0	23
3	13	63	0	25
2	33	17	33	17
1	0	20	20	60
Total Means	7.4	8.0	4.8	6.0

perhaps explained by the probability that persons unable to match a straight-forward iambic pentameter line, one of the standard English verse forms, with similar surrounding lines, are operating at a chance level. The item was fairly successful, however, in determining which students are able to detect a reasonably straightforward metrical pattern.

(3) Rhyme

Conventional end rhyme is probably the most commonly used and easily detected kind of sound repetition found in English poetry. It makes an important contribution to the musical effect as well as the structure of a poem, less essentially but more obviously than metre and rhythm. Because of this obviousness it was concluded that "full rhyme" items would fail to distinguish between the more and less able poetry readers and hence would be of limited usefulness for the present test. However, one such item (A-1) was included in order to test this assumption.

Half-rhyme, or slant-rhyme, as opposed to full rhyme, is rarely used and could be expected to be difficult to detect. To test the degree of difficulty, two items measuring response to half-rhyme were included in the Poetry Test. Thus the Poetry Test includes three items intended to measure response to rhyme. Two of these, one based on Dickinson's "Railway Train" and one based on Owen's "Arms and the Boy", dealt with half, or slant-rhyme. As might be expected, a detailed examination of item data for these two

items reveals a low level of awareness of half rhyme in the group as a whole, with much higher figures for the Grade XII group than for the Grade X group. Grade X students, it appears, have little or no awareness of half rhyme. The irregularity of the data patterns for these two items suggests that for the most part selection among the alternatives was not made on the basis of a knowledge of rhyme.

Consequently, the third item dealing with rhyme, taken from a little known poem by Auden, was selected for detailed examination. The results of the item, which deals with full rhyme, are presented in Table XXXI.

"Paid on Both Sides"

Table XXXI, which presents item data for Item 1, Form A, reveals that in constructing a full-rhyme item for such a test as the present one, it is almost impossible to avoid producing one which is overly easy. It may be noted, however, that with two of the five groups to whom the item was administered, more than 25 per cent of the respondents failed to detect the rhyme. Although it seems reasonable to assume that in an item as easy as this most students failing to detect the correct answer will be operating at the chance level, an examination of the data for the pilot study and the final test administration shows "hills", the first alternative, as a consistent second choice and "river", the second alternative, as a consistent last choice. This can perhaps be explained by the attractive

TABLE XXXI

Detailed Item Data for Item A-1¹

Though he believe it, no man is strong,
 He thinks to be called the fortunate,
 To bring home a wife, to live long.

But he is defeated; let the son
 Sell the farm lest the mountain fall;
 His mother and her mother won.

His fields are used up where the moles visit,
 The contours worn flat; if there show
 Passage for water he will miss it:

Give up his breath, his woman, his team;
 No life touch, though later there be
 Big fruit, eagles above the _____.

1. hills
2. river
3. sea
4. stream*

Series 2 Difficulty .86 Biserial Corr. .33

Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	2	6	6	86
5	0	0	8	92
4	0	0	13	88
3	0	0	0	100
2	8	17	0	75
1	0	9	9	82
Total Means	7.0	6.7	8.0	9.0

¹W. H. Auden, Chorus from "Paid on Both Sides," in
Selected Poems (London: Faber and Faber, 1938), p. 38.

TABLE XXXI (Continued)

Series X	Difficulty .74		Biserial Corr. .49	
Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	2	11	11	75
5	0	0	0	100
4	0	11	11	78
3	0	50	0	25
2	0	0	15	85
1	13	25	25	38
Total Means	6.0	10.0	9.6	12.4

Pilot	Difficulty .87		Biserial Corr. .38	
Total	9	0	3	88
5	0	0	3	100
4	25	0	0	75
3	9	0	0	91
2	0	0	0	100
1	14	0	0	7
Total Means	9.0	-	8.0	10.5

Final - Total Group	Difficulty .76		Biserial Corr. .49	
Total	19	1	4	76
5	0	0	0	100
4	7	0	7	86
3	40	0	0	60
2	29	0	0	72
1	29	4	8	58
Total Means	4.5	4.0	5.8	6.6

TABLE XXXI (Continued)

Final - Grade X Difficulty .73 Biserial Corr. 150

Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	20	2	4	73
5	0	0	17	83
4	8	0	0	92
3	25	0	0	75
2	17	8	8	67
1	67	0	0	33
Total Means	3.6	4.0	6.0	5.9

Final - Grade XII Difficulty .80 Biserial Corr. .36

Total	17	0	2	80
5	0	0	0	100
4	8	0	8	85
3	25	0	0	75
2	33	0	0	67
1	20	0	0	80
Total Means	5.6	-	8.0	7.4

flavour of "hills" and the fact that "river" is a two-syllable word. More significant, possibly, is a comparison of test means for respondents selecting alternatives 2 and 3. These means are consistently high for "sea", the third alternative, and low for "river", the second alternative. Although both these alternatives were unpopular choices, "sea", unlike "river", was selected by the more able students. A possible explanation for this selection is that "sea" and "stream", the correct answer, are rhymes of a sort.

This item, then, was reasonably successful in its overall discriminating power and in its ability to separate respondents aware of ordinary rhyme from those who are not. The item also indicates little difference between the ability of Grade X and Grade XII students to detect a conventional rhyming pattern.

(4) Metaphor

(a) "Bereft"

Table XXXII presents detailed item data for Item 11, Form B, based on Frost's poem "Bereft". In this poem the speaker, projecting his fear and feeling of desolation into his view of his natural surroundings which he describes in sinister terms, uses an extended metaphor in which a leaf-whirlwind is compared with a snake. By deleting one part of this metaphor and supplying distractors which fail to complete it, an attempt was made to measure students' responsiveness to the metaphoric or figurative aspect of

TABLE XXXII

Detailed Item Data for Item B-11¹

Where had I heard this wind before
 Change like this to a deeper roar?
 What would it take my standing there for,
 Holding open a restive door,
 Looking down hill to a frothy shore?
 Summer was past and day was past,
 Sombre clouds in the west were massed.
 Out in the porch's sagging floor,
 Leaves got up in a _____ and hissed,
 Blinding struck at my knee and missed.
 Something sinister in the tone
 Told me my secret must be known:
 Word I was in the house alone
 Somehow must have gotten abroad,
 Word I had no one left but God.

1. whirl
2. coil*
3. ring
4. twist

Series 1	Difficulty .46		Biserial Corr. .51	
Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	35	46	2	17
5	31	69	0	0
4	25	63	0	12
3	20	60	0	20
2	44	22	12	22
1	45	19	0	36
Total Means	6.8	8.6	6.0	5.6

¹Robert Frost, "Bereft," in A Little Treasury of Modern Poetry, ed. Oscar Williams (New York: Scribner's, 1952), p. 138.

TABLE XXXII (Continued)

Series X	Difficulty .59		Biserial Corr. .58	
Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	5	59	9	27
5	0	100	0	0
4	11	56	0	33
3	0	75	25	0
2	9	39	15	38
1	0	38	12	50
Total Means	11.0	13.0	9.5	9.5

Pilot	Difficulty .47		Biserial Corr. .35	
Total	43	47	0	10
5	14	86	0	0
4	67	33	0	0
3	40	30	0	30
2	67	33	0	0
1	57	43	0	0
Total Means	8.1	10.4	-	9.3

Final - Total Group	Difficulty .36		Biserial Corr. .67	
Total	40	36	4	20
5	0	76	12	12
4	29	50	0	21
3	31	38	0	31
2	61	26	0	13
1	59	0	12	29
Total Means	5.4	8.9	6.5	6.3

TABLE XXXIII (Continued)

Final - Grade X Difficulty .26 Biserial Corr. .48

Selecting Group	% Choosing Alternatives			
	1	2*	3	4
Total	47	27	2	24
5	13	50	0	37
4	25	38	0	37
3	57	29	0	39
2	63	25	0	12
1	64	0	9	27
Total Means	5.0	7.0	2.0	6.3

Final - Grade XII Difficulty .45 Biserial Corr. .71

Total	34	45	6	15
5	0	89	11	0
4	17	76	8	8
3	45	27	0	28
2	86	14	0	0
1	38	12	12	38
Total Means	6.3	10.0	8.0	6.3

poetic language.

As the figures for the final test administration show, the item appears to be highly successful in this regard, particularly with the Grade XII group, although it proved to be on the difficult side for the Grade X students. By far the most popular distractor in all test administrations proved to be "whirl". This distractor made its greatest appeal to the less able students, probably because of its association with "whirlwind" but perhaps because, metaphoric considerations aside, it is a better choice than "circle" and "ring", which are visually less accurate, and "twist", which forms an annoying internal rhyme. It might be noted that "circle", which was substituted for "whirl" in the second series as a less tempting distractor, proved to be far less appealing, probably for both metric and descriptive reasons. Alternative 3, "circle", was consistently the least popular of the four, probably because of its visual inaccuracy, but possibly for connotative reasons. "Twist", the fourth alternative, proved to be consistently more popular than "ring" and "circle", probably because of its greater visual accuracy and suitably sinister connotation. Generally only the less able students were attracted by "twist". Apparently the abler respondents were deterred by the inappropriate internal rhyming of "twist" and "hissed".

This item appeared to achieve its main objective of determining which students could complete the metaphor. In

addition it supplied some incidental information. As with the tone and metre items discussed above, the Grade XII group achieved almost twice as high a score on the item as was achieved by the Grade X group. The one possible weakness of the item is that the third alternative, "ring", was selected by so few students that it is barely operative as a distractor. However, variation in the attractiveness of the distractors, unless too pronounced, was not considered a serious limitation for the purposes of the present test. Indeed, such variability, in differentiating between students with sufficient knowledge of the principle being measured to select a plausible distractor and students selecting more obviously inappropriate distractors, can supply more information than would be obtained by the use of distractors of uniform attractiveness.

(5) "Thought", or Plain Sense Meaning

Eight of the 32 items of the Poetry Test sought to measure the degree to which a student grasped some aspect of the "thought", or paraphrasable content, of a poem. These items varied considerably in the exact skill they were attempting to measure. Some, like the ones based on "In the Restaurant," which tried to test a student's skill in making an inference from a dramatic situation, demanded a high degree of awareness of the whole poem. Others, like the item based on "Central Park Tourney", involved several contextual hints scattered throughout the poem.

(a) "In the Restaurant"

Table XXXIII presents detailed item data for Item 13, Form B, based on Hardy's "In the Restaurant". In constructing this item, an attempt was made to ensure that a full awareness of the dramatic situation developed by the poem as a whole would be required in making the correct selection. An attempt was also made to construct one distractor, alternative 1, to attract respondents looking only at the context of the last two lines; one distractor, alternative 2, which conflicts with the context of the following line as well as the context of the rest of the poem; and one distractor, alternative 3, which is patently self-contradictory. Thus, it was reasoned that alternative 3 would tempt only the least wary and alternative 2 those respondents only slightly less observant, while alternative 1 would act as the most effective distractor.

An examination of Table XXXIII shows that the results, in terms of numbers of students selecting each alternative, reflect the intended strategy. Of the three distractors, "stay and hide it" is consistently the least popular. Indeed, "stay and face it" proved to be more popular than the original for two of the four groups to which the item was administered and served well in discriminating between students who understood the whole poem and those who did not.

It might be noted that for this item the Grade XII scores were more than twice as high as the Grade X scores,

TABLE XXXIII

Detailed Item Data for Item B-13¹

But hear. If you stay, and the child be born,
 It will pass as your husband's with the rest,
 While, if we fly, the teeth of scorn
 Will be gleaming at us from east to west;
 And the child will come as a life despised;
 I feel an elopement is ill-advised!

O you realize not what it is my dear,
 To a woman! Daily and hourly alarms
 Lest the truth should out. How can I stay here
 And nightly take him into my arms!
 Come to the child no name or fame,
 Let us _____, and bear the shame.

1. stay, and face it
2. go, and hide it
3. stay, and hide it
4. go, and face it*

Series 2 Difficulty .31 Biserial Corr. .42

Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	47	12	10	31
5	31	8	8	54
4	25	0	25	50
3	57	14	0	29
2	67	17	8	8
1	55	18	9	18
Total Means	8.3	7.7	8.2	10.0

¹Thomas Hardy, "In the Restaurant," in Modern British Poetry, ed. Louis Untermeyer (New York: Harcourt, Brace, and World, 1962), p. 30.

TABLE XXXIII (Continued)

Pilot	Difficulty .43		Biserial Corr. .34	
Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	33	20	3	43
5	29	29	0	43
4	0	0	0	100
3	40	20	0	40
2	0	33	0	67
1	57	14	14	14
Total Means	9.0	9.5	5.0	10.4

Final - Total Group	Difficulty .36		Biserial Corr. .54	
Total	38	17	9	36
5	12	6	12	71
4	29	21	7	43
3	44	13	13	31
2	42	23	6	29
1	59	18	12	12
Total Means	5.7	6.3	6.6	8.5

TABLE XXXIII (Continued)

Final - Grade X	Difficulty .24		Biserial Corr. .37	
Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	47	20	8	24
5	38	13	13	38
4	50	13	13	25
3	43	21	0	36
2	25	25	25	25
1	73	27	0	0
Total Means	5.3	5.4	6.8	6.8

Final - Grade XII	Difficulty .49		Biserial Corr. .54	
Total	28	13	11	49
5	11	0	0	89
4	8	25	17	50
3	36	9	9	45
2	57	14	0	29
1	38	12	25	25
Total Means	6.4	7.8	6.4	9.4

indicating a dramatic increase in the ability to make the kind of inference demanded by this item with grade level. The required kind of inference seems to depend on sensitivity to human experience rather than on sensitivity to language per se.

(b) "Central Park Tourney"

Table XXXIV presents item data for Form B, Item 3, taken from Mildred Weston's "Central Park Tourney", in which the visual resemblance of opposing car lights and jousting knights suggests continuing human competitiveness and restless opposition. The item itself, however, does not deal with the deeper meanings of the poem: it merely seeks to measure the student's ability to infer the identity of the first element of a metaphoric comparison from knowledge of the second element. This item was intended as an easy one, and the results presented in Table XXXIV show that it was successful in this respect. Of the various distractors tried out during the successive administrations of the item, "men" and "boys" understandably attracted very few of the respondents. "Horses" proved unsuccessful as a distractor, attracting only one or two of the least able students, undoubtedly because its selection would involve a nonsensical comparison of horses with horses. "Steeds" also proved unattractive, although slightly more tempting than "horses", possibly because, illogical as it appears as a choice, it better fits the metrical pattern of the passage and connotes

TABLE XXXIV

Detailed Item Data for Item B-3¹

In the Park	Staged
With long spear lights	In the Park
Ride at each other	From dusk
Like armoured knights;	To dawn,
Rush,	The tourney goes on:
Miss the mark,	Rush,
Pierce the dark,	Miss the mark,
Dash by!	Pierce the dark,
Another two	Dash by!
Try.	Another two
	Try.

1. Bikes
2. Boys
3. Cars*
4. Steeds

Series 1 Difficulty .80 Biserial Corr. .32

Selecting Group	% Choosing Alternatives			
	1	2	3*	4
Total	2	15	80	2
5	0	15	85	0
4	0	13	88	0
3	0	0	100	0
2	0	22	78	0
1	0	18	64	9
Total Means	5.0	6.6	7.5	5.0

¹Mildred Weston, "Central Park Tourney," in Poetry, ed. K. P. Dover (Toronto: Holt Rinehart and Winston, 1964), p. 7.

TABLE XXXIV (Continued)

Series X Difficulty .57 Biserial Corr. .36

Selecting Group	% Choosing Alternatives			
	1	2	3*	4 ~~≠~~
Total	2	5	57	36
5	0	0	90	10
4	11	11	33	44
3	0	0	25	75
2	0	8	69	23
1	0	0	38	63
Total Means	14.0	10.5	12.5	10.3

* Changed for Series X only from "Steeds" to "shadows").

Pilot Difficulty .57 Biserial Corr. .43

Total	33	7	57	3
5	14	0	86	0
4	0	0	67	33
3	50	10	40	0
2	0	0	100	0
1	57	14	29	0
Total Means	8.6	7.0	10.5	11.0

Final - Total Group Difficulty .67 Biserial Corr. .15

Total	17	7	67	8
5	12	0	88	0
4	0	14	86	0
3	13	6	75	6
2	13	10	68	10
1	47	6	24	24
Total Means	5.3	6.0	7.6	4.8

TABLE XXXIV (Continued)

Final - Grade X	Difficulty .61		Biserial Corr. .46	
Selecting Group	% Chcoosing Alternatives			
	1	2	3*	4
Total	20	8	61	10
5	0	0	100	0
4	13	13	74	0
3	21	7	57	14
2	13	25	50	12
1	45	0	37	18
Total Means	4.6	5.8	6.4	4.6

Final - Grade XII	Difficulty .74		Biserial Corr. .52	
Total	13	6	74	6
5	11	0	89	0
4	8	8	83	0
3	9	9	73	9
2	0	0	100	0
1	38	13	25	25
Total Means	6.3	6.3	8.8	5.0

the age of chivalry.

It might be noted that "shadows", nonsensical as it is in the context of the poem from a denotative standpoint, received a surprisingly large number of votes, especially from the less able respondents in the second set of data. This outcome might be explained by the probability that "shadows" is an attractively nebulous word, highly evocative, whose connotations do not run counter to the prevailing tide of the poem and which students have been conditioned to regard as "poetic".

"Bikes", which turned out to be the most consistently attractive distractor, was constructed for this purpose, and appears conceptually to be the most likely alternative to "cars". It seems likely that the students successful on this item used clues such as "like armoured knights" and the uninterrupted sequence of the near-collisions to eliminate "bikes".

Unlike the previous item this one proved about equally difficult for Grade X and XII students. This suggests that this item is operating as a considerably less sophisticated level of inference than "Restaurant", the one previously discussed.

(6) Word Flavour

As was noted in Chapter II, most current critical opinion agrees that sound values in a word or group of words, even in the case of direct onomatopoeia, where word sounds

are meant to parallel rather than merely suggest actual sounds ("caw", "clip-clop", and "meow" are common examples) act not in isolation but in conjunction with the other meanings, mainly connotative, which the word or expression may have. Thus in the present test it was found to be impossible to measure with any accuracy a student's response to sound alone. The title of this section, "Word Flavour", reflects this inability and is meant to suggest a mixture of sound and suggestion.

In the construction of the ten items which could be described as intended primarily to measure "word flavour", two general tendencies became apparent. First, it was noted that even the slightest changes in the associational aspect of the distractors resulted in substantial changes in response patterns, and completely overshadowed changes in the sound qualities of the words involved. Thus, in Item 6, Form A, based on Milton's famous sonnet beginning "Avenge, O Lord, thy slaughtered saints," an attempt to measure students' ability to match the various alternatives to the content in terms of Milton's rather distinctive "sound"--his wrathful organ tones--was partially thwarted by the likelihood that selection among the alternatives was based more on their associational quality than on their sound. Thus "fallen" was preferred to "murdered", which in turn was preferred to "butchered", although on the basis of sound values alone "butchered" matches the context more closely and "fallen" less closely. This tendency is also

noted in the two items examined below, one based on an excerpt from Julius Caesar and the other on the deservedly famous poem "Study Hall".

A second tendency noted in the construction of "word favour" items is that the more the item measures the sound quality, as opposed to the suggestiveness of a word, the more the Grade X results resemble the Grade XII results in terms of difficulty level. This tendency suggests that a sensitivity to the sounds of words tends to level off earlier than most of the other skills which the test attempts to measure. Sensitivity to the sounds of words may be a more basic and unconscious phenomenon than sensitivity to the other kinds of meanings, both denotative and connotative, which are more dependent on consciously acquired experience. It may be more a matter of temperamental differences than of formal learning. If this conclusion is warranted--and this matter clearly bears further investigation--it should be noted by the teacher of poetry.

(a) from Julius Caesar

Table XXXV presents item data for Item 6, Form B.

In the construction of the distractors for this item, every effort was made to keep constant all aspects of poetic meaning except sound. There was no problem in producing alternatives satisfactorily equivalent denotatively and metrically. The real challenge was to control the connotative element. It was felt that this control was finally

TABLE XXXV

Detailed Item Data for Item B-6¹

O, pardon me, thou bleeding piece of earth,
 That I am meek and gentle with these butchers!
 Thou art the ruins of the noblest man
 That ever lived in the tide of the times.
 Woe to the hand that shed this costly blood!
 Over thy wounds now do I prophesy
 (Which, like dumb mouths do ope their ruby lips
 To beg the voice and utterance of my tongue),
 A curse shall light upon the limbs of men;
 Domestic fury and
 Shall cumber all the parts of Italy;
 Blood and destruction shall be so in use
 And dreadful objects so familiar
 That mothers shall but smile when they behold
 Their infants quartered with the hands of war,
 All pity choked with custom of fell deeds.

1. fierce civil war
2. wild civil strife
3. fierce violence
4. fierce civil strife*

Series 2 Difficulty .39 Biseriail Corr. .48

Selecting Groups	% Choosing Alternatives			
	1	2	3	4*
Total	21	25	13	41
5	0	25	8	67
4	43	14	14	29
3	20	50	0	30
2	27	9	18	46
1	25	25	28	17
Total Means	8.3	9.8	8.1	10.9

¹Wm. Shakespeare, *Julius Caesar* (Act III, Scene i),
 ed. G. L. Kittredge (Boston: Ginn, 1939), p. 50.

TABLE XXXV (Continued)

Pilot	Difficulty .47		Biserial Corr. .38	
Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	20	30	3	47
5	14	14	0	72
4	0	0	0	100
3	40	30	10	20
2	0	67	0	33
1	14	43	0	43
Total Means	8.8	8.7	10.0	10.4

Final - Total Group	Difficulty .49		Biserial Corr. .26	
Total	13	29	9	49
5	12	12	12	64
4	21	21	14	44
3	13	25	12	50
2	10	32	3	55
1	6	53	12	29
Total Means	7.3	5.6	7.2	7.5

TABLE XXXV (Continued)

Final - Grade X Difficulty .45 Biserial Corr. .12

Selecting Group	% Choosing Alternatives			
	1	2	3	4*
Total	16	33	6	45
5	25	13	12	50
4	25	13	12	50
3	14	43	7	36
2	13	38	0	50
1	9	45	0	45
Total Means	6.6	4.8	7.0	6.0

Final - Grade XII Difficulty .55 Biserial Corr. .33

Total	6	26	13	55
5	0	11	11	78
4	17	17	17	50
3	9	36	9	45
2	0	14	0	86
1	0	50	25	25
Total Means	9.0	6.6	7.3	8.8

successfully achieved: the emotional colouring of, say, "fierce violence" and "wild civil strife" appeared to be very nearly the same and to match the context of the excerpt as well as of the original phrase "fierce civil strife". The only variable appeared to be the sound, in which the four alternatives, by means of various arrangements of consonant clusters, ranged from the extreme cacophony and explosiveness of the original, "fierce civil strife", through "wild civil strife" and "fierce civil war" to the relatively mellifluous "fierce violence".

An examination of Table XXXV shows that the item was generally successful in separating students more sensitive to the sound qualities of a group of words from those less sensitive to sound qualities. Also, that alternative 3, considered the poorest match, proved to be consistently the least attractive choice indicated that the item was successful in modulating the degree of appeal of the three distractors, and hence, presumably, in identifying the respondents least sensitive to word sounds.

A glance at the biserial correlation coefficients and the test means of the students responding to the various alternatives, however, indicates either that the item has faulty discriminating ability or that it measures a skill very different from the composite skill measured by the test as a whole. That the test means for the distractor constructed as the poorest choice are, if anything, higher than those of the other distractors gives some credence to

to the first of these explanations, that the item in general is deficient in discriminatory power. In view of the difficulties in measuring sound values as separate from connotation, and in view of the subtle differences between the four alternatives on the sound dimension, this lack of discriminating ability is perhaps not unexpected.

It might be of interest, finally, to compare the item data for the Grade X and Grade XII groups. The overall difficulty level is quite similar for the two groups, but while it is quite clear that for the Grade XII group the "correct" answer attracted the abler students in terms of total test performance, for Grade X there is virtually no correspondence between success on this item and success on the test as a whole. Perhaps this serves to confirm an earlier remark about responsiveness to word sounds as a largely unconscious phenomenon not necessarily related to other kinds of word meanings. This matter appears to warrant further investigation, and is of considerable import not only for response to poetic language but for response to imaginative language as a whole.

(b) "Study Hall"

Table XXXVI presents item data for Item 16, Form B. This item was selected for detailed examination largely because having been presented in four different forms to six different groups of students it promises, perhaps better than any of the items discussed previously, to

TABLE XXXVI

Detailed Item Data for Item B-16¹

Smuggled sibilants,
 Rattling pages,
 A muted yawn,
 Someone _____ ages away,
 Faint footsteps overhead,
 A clicked looseleaf,
 Desultory drumming of bored pencils;
 A faintly sweaty, chalky smell;
 A weary, yet taut, expectation
 Of the noon bell.

1. droning*
2. whispering
3. laughing
4. muttering

Series 1 Difficulty .67 Biserial Corr. .32

Selecting Group	% Choosing Alternatives			
	1*	2	3 [‡]	4 [‡]
Total	67	27	3	3
5	86	14	0	0
4	100	0	0	0
3	67	23	0	0
2	50	25	13	13
1	57	43	0	0
Total Means	7.3	6.0	6.0	6.0

[‡] Changed for Series 1 only from "laughing" and "muttering" to "talking" and "speaking".

¹Xavier Warden, "Study Hall." Unpublished poem, 1958.

TABLE XXXVI (Continued)

Series 2 Difficulty .53 Biserial Corr. .50

Selecting Group	% Choosing Alternatives			
	1*	2	3 [≠]	4 [≠]
Total	53	41	2	4
5	77	23	0	0
4	75	13	0	13
3	43	57	0	0
2	42	42	8	8
1	27	73	0	0
Total Means	9.7	7.5	8.5	8.5

[≠] Changed for Series 2 only from "laughing" and "muttering" to "jabbering" and "babbling".

Series X Difficulty .36 Biserial Corr. .20

Total	36	25	32	7
5	40	30	20	10
4	44	44	11	20
3	50	0	25	25
2	31	15	54	10
1	25	25	38	12
Total Means	12.4	11.5	10.9	11.3

Pilot Difficulty .63 Biserial Corr. .61

Total	63	17	3	17
5	86	0	3	11
4	67	19	14	0
3	80	0	0	20
2	33	33	0	34
1	29	43	0	28
Total Means	10.5	7.2	12.0	7.8

TABLE XXXVI (Continued)

Final - Total Group Difficulty .41 Biserial Corr. .47

Selecting Group	% Choosing Alternatives			
	1*	2	3	4
Total	39	34	7	16
5	71	18	0	6
4	50	36	7	7
3	50	25	0	25
2	16	45	23	13
1	29	35	12	24
Total Means	8.2	6.2	5.7	6.3

Final - Grade X Difficulty .39 Biserial Corr. .42

Total	37	33	6	18
5	75	13	0	12
4	62	13	0	25
3	14	50	14	22
2	13	50	13	14
1	36	27	0	27
Total Means	6.7	5.3	5.7	5.2

Final - Grade XII Difficulty .43 Biserial Corr. .55

Total	43	34	9	13
5	78	11	0	11
4	50	33	0	17
3	36	46	0	18
2	29	29	42	0
1	13	50	13	14
Total Means	9.6	7.1	5.8	8.0

reveal the effects of changes in item form on test results.

In its final form, as administered during the pilot study and in the final administration, the item was generally successful and attained suitable difficulty levels and relatively high biserial correlation coefficients. As in most of the items in which word sound played a prominent part, the Grade X and Grade XII difficulty levels were similar. The probable reason for this has been discussed.

In the initial construction and subsequent revisions of the item an attempt was made to provide distractors which suited the context equally as well as the original version on the denotative and metrical dimensions, but which varied in their suitability to the context according to suggestive and sound qualities. Thus it was felt that while all of the distractors make equal sense in terms of the context of the whole poem, "droning", the original, in its connotation and sound best suits the atmosphere of boredom the poem seeks to convey.

From the item data based on the initial form of the item, it appeared that the item had promise--that "whispering", the second alternative, was a successful distractor in terms of attractiveness and discriminating power, but that the overall value of the item was weakened by the lack of attractiveness and hence of effectiveness of the third and fourth alternatives, "talking" and "speaking". Consequently these were changed, again without sufficient attention to potential attractiveness, to "jabbering" and

"babbling". As might be expected these were both rejected by the vast majority of the respondents. It might be noted that "talking" and "speaking" were probably rejected because of a lack of evocative and sound colour, while "jabbering" and "babbling", although highly coloured words, were probably rejected because their association and tonal colouring is too obviously at odds with the rest of the poem. "Shouting" and "laughing", in the second and third revisions, seem to be in much the same position as "talking" and "speaking", i.e., they are too neutral connotatively. These distractors do, however, appear to possess more connotative intensity than "speaking" and "talking", although to some extent their overtones, like those of "jabbering" and "babbling", run too obviously counter to the prevailing mood of the poem. Again here is a rather subtle matter which requires systematic and detailed investigation.

It could be argued that "muttering", in terms of the atmosphere of the whole poem, is the second best choice, slightly more suitable than "whispering". It turned out to be a reasonably attractive distractor, and in the final administration of the item tended to attract slightly more able students than did "whispering".

Thus after many changes the item turned out to be successful not only in terms of difficulty level and overall discriminating ability but in producing a response which reflected reasonably accurately the conceptual framework upon which it was constructed. In addition to identifying

both the most able and least able students in terms of total test performance the item provides information upon which inferences regarding the selection of one kind of word or another could be based. From these inferences it is hoped that sharper, more effective items of this general nature can be constructed.

Summary of the discussion of individual items. The foregoing relatively detailed examination of certain items was intended chiefly to indicate the kinds of information regarding student response to poetic language which a test of this kind can provide. Because of the largely exploratory nature of the study, the experimental test forms were intended more to give direction to further, more rigorously controlled investigation than to provide conclusive information. Nevertheless certain generalizations emerging from the previous discussion can be presented in summary form:

(1) General Comments

a. The type of testing procedure used in the present study appears to be effective in measuring a wide variety of the component skills involved in deriving meaning from a poem. These skills include awareness of tone, metrical pattern, various types of rhyme, the operations of figurative language, the effect of word sounds, and the implied dramatic situation in a poem.

b. The present testing procedure also appears to make possible the construction of items which measure these various skills at different levels of sophistication by including distractors either grossly or subtly inappropriate to the context.

c. Apparently minor changes in the alternatives provided for an item often produced fundamental changes in response patterns. This confirms the view widely held by literary critics that the relationship, or balance, between a word or expression in a poem and its context is an extremely delicate one, and makes the construction and refinement of items such as those used in the present test an exacting task.

d. An item administered unchanged to different groups can produce widely differing response patterns which are not necessarily due to grade level. This mercurial quality of response for some items is potentially a serious weakness of the present kind of testing and must be remedied in order to obtain a satisfactorily reliable and valid measuring device.

e. As could be expected, the number of students selecting a given alternative is not necessarily related to the kind of student making the selection. Clearly, both kinds of data need to be carefully examined in assessing and interpreting an item. Certain items apparently too easy or too difficult in terms of the 0.50 difficulty level criterion endorsed by most testing authorities are

considered useful for the present testing procedure if they possess adequate discriminating power. These items can be effective in isolating students deficient in "easy" skills and those showing mastery of "difficult" skills.

(2) Comments Regarding Specific Skills

a. Grade X students were generally unable to respond adequately to the "tone" items.

b. In response to the metrical items, line length appeared to be more important than smoothness of metrical flow.

c. Contrary to expectation, the metre (rhythm) items were generally no easier than the other kinds of items. This matter will be elaborated upon in Chapter V in the discussion of the interview results.

d. Full rhyme, generally, was easily detected. Even so, approximately one quarter of the students in the total sample were unaware of the rhyme in the one "conventional rhyme" item.

e. Half rhyme, or slant rhyme, measured in two of the 32 items, was detected by less than a quarter of the respondents.

f. In responding to items dealing with connotative meaning or relatively subtle denotative meaning the less able students appeared to be influenced more by the quality of the alternative itself (its "poetic" appeal, its internal consistency, its "stock" association) than by its

appropriateness to its context. Cases in point are "whirl" (page 172), "shadows" (page 182), and "stay and face it" (page 174).

g. Grade level appeared to be a highly important factor in response to items involving an implied, rather than openly stated, plot or dramatic situation.

h. In attempting to measure sensitivity to the sound value of a word or expression, it was found almost impossible to vary the "sound" without varying the suggestive quality even more (page 183). This supports the generally held critical view that word sounds in poetry operate only in conjunction with other kinds of meaning. It also makes the construction of purely "sound" items for a test of this nature highly difficult.

i. For the item which appeared most successful in measuring "sound" as distinct from "suggestion" (B-6), there is little difficulty-level difference between Grade X and Grade XII students. This indicates a certain levelling off of sensitivity to the purely "sound" aspect of poetic meaning after Grade X (page 184).

In Chapter VI, some of the implications for further research suggested by the above findings will be discussed.

PERFORMANCE ON THE POETRY TEST
BY VARIOUS GROUPS

Grade Level Differences

Hypothesis 1 on page 13 states:

There is no significant difference between the mean performance of Grade X students and the mean performance of Grade XII students on the Poetry Test.

This hypothesis was tested by means of two-way analyses of variance¹ for (a) Grade Level by Test Form, (b) Grade Level by Sex, and (c) Grade Level by Reading Test performance. Grade Level effects were used as a common element in these analyses in order to facilitate comparison of the relationships between Grade Level, Sex, and Reading Level and Poetry Test performance.

Because of the nature of the data a procedure assuming a fixed effects model and allowing for unequal numbers of observations per cell was selected.

The complete analysis-of-variance data for each of the subtests appear in Appendix F. The four tables presented here deal only with total Poetry Test scores. Table XXXVII presents the results for the analysis of variance for Test Form by Grade Level. Grade Level Main Effects are significant at the 0.001 level of confidence, while Test Form Main Effects are not significant at the 0.05 level. There is no significant interaction between the two factors. However, Table XXXVII also shows a serious lack of homogeneity of the population error variances in this analysis.

¹G. A. Ferguson, Statistical Analysis (New York: McGraw-Hill, 1966), pp. 300-323.

Because homogeneity of variance is one of the assumptions underlying the F statistic, this lack places the accuracy of the decision reached by the F test in question, and dictates that these data be interpreted with caution. Nevertheless, Winer¹ observes that moderate departures from homogeneity of variance do not seriously affect the accuracy of these decisions. Kirk² further notes that such departures are not too serious if the number of subjects in each cell is the same. Since these cells are very nearly equal in the present analysis and the Grade Level effects are highly significant ($p < 0.001$) these data appear to support the conclusions, previously reported, that the forms are not parallel and that the Grade X students responded at close to the chance level.

TABLE XXXVII

Two-Way Analysis of Variance on the Poetry Test
 Factor A. Test Form
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	p
"A" Main Effects	19.8	1	19.8	3.43	.07
"B" Main Effects	195.4	1	195.4	33.81	< .001
"AB" Interaction	5.5	1	5.5	.95	.33
Error Term	1074.7	186	5.8		
Homogeneity of Variance Test	Chi square = 14.50 p = <.001				

¹B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw-Hill, 1962), p. 33.

²Roger Kirk, Experimental Design in the Behavioral Sciences (Belmont, Cal.: Brooks and Cole, 1968), pp. 61-62.

Table XXXVIII presents results for the two-way analysis of variance for Sex and Grade Level. For both Forms, A and B, Grade Level Main Effects are significant at the .01 level. Sex differences in performance are not statistically significant and there is no significant interaction.

Table XXXIX presents results for the two-way analysis of variance of the Poetry Test criterion scores for Grade Level and Reading Level, as determined by Reading Test Totals. For Form A, Grade Level Main Effects are significant at the .01 level, while Reading Level Main Effects do not attain statistical significance, and there is no significant degree of interaction. For Form B, Grade Level Main Effects are significant at the .001 level, Reading Level Main Effects at the .05 level, and there is no significant degree of interaction.

The results presented in these three tables indicate that Grade Level is more significant than Reading Level and Test Form, and substantially more significant than Sex, in contributing to variance in the Poetry Test total scores. Null hypothesis 1, therefore, is rejected.

It appears reasonable that this substantial Grade-Level influence is due to such factors as maturation, practice in communication skills, the study of literature in general, and the study of poetry in particular. However, determining the relative importance of such factors is not within the scope of the present study.

TABLE XXXVIII

Two-Way Analysis of Variance
 Factor A. Sex
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	p
<u>FORM A</u>					
"A" Main Effects	1.3	1	1.3	.29	.59
"B" Main Effects	62.0	1	62.0	13.74	< .001
"AB" Interaction	3.0	1	3.0	.65	.42
Error Term	401.7	89	4.5		
Homogeneity of Variance Test	Chi square = 3.52			p = .32	
<u>FORM B</u>					
"A" Main Effects	3.7	1	3.7	.58	.45
"B" Main Effects	141.7	1	141.7	22.02	< .001
"AB" Interaction	5.7	1	5.7	.88	.35
Error Term	591.7	92	6.4		
Homogeneity of Variance Test	Chi square = 5.31			p = .15	

TABLE XXXIX

Two-Way Analysis of Variance
 Factor A. Reading Level
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	p
<u>FORM A</u>					
"A" Main Effects	23.4	2	11.7	2.71	.07
"B" Main Effects	63.0	1	63.0	14.58	< .001
"AB" Interaction	20.0	2	10.0	2.31	.11
Error Term	319.8	74			
Homogeneity of Variance Test				Chi square = .48	p = .44
<u>FORM B</u>					
"A" Main Effects	48.6	2	24.3	4.05	.02
"B" Main Effects	93.2	1	93.2	15.54	< .001
"AB" Interaction	19.7	2	9.8	1.64	.20
Error Term	473.5	79	6.0		
Homogeneity of Variance Test				Chi square = .10	p = .07

As a further means of testing Hypothesis 1, means and standard deviations, grouped by grade level, are presented in Table XL. It might be noted that, overall, there is a mean increase of approximately 37 per cent in Poetry Test totals of Grade XII scores over Grade X scores, while in Reading Test totals there is a mean increase of approximately 30 per cent. Looking at the subtest means, it appears that overall mean increases for Subtests One (Thought) and Three (Sound) are also approximately 30 per cent, paralleling the Reading Test increases. Mean increases for Subtests Two (Suggestion) and Four (Rhythm), of approximately 50 and 45 per cent respectively, are considerably higher. A likely explanation for this is that Subtests One and Three measure skills which, like those measured by the Reading Test, tend to fall in the category of "general language ability", or "general ability in communication". Subtests Two and Four, on the other hand, may measure skills more idiosyncratic to poetry, perhaps more subject to rapid increases as a result of specialized instruction. If this explanation is valid, the implication for the teacher of English is that, although some of the components of skill in responding to a poem develop more or less naturally, others must be intensively and artificially cultivated. Hence, both extensive and intensive approaches to the teaching of poetry appear necessary, with each complementing the other.

TABLE XL
Means and Standard Deviations of Poetry Test Totals
and Subtotals Compared by Grade Level

Poetry Scores	Form A			Form B			Mean % Increase
	Gr. X	Gr. XII	Per cent Increase	Gr. X	Gr. XII	Per cent Increase	
Total							
Mean	5.47	7.16	31	5.78	8.22	42	37
S.D.	1.87	2.31		2.05	2.87		
Subtest 1							
Mean	1.55	2.11	36	2.00	2.41	20	28
S.D.	.95	.93		.96	.90		
Subtest 2							
Mean	1.57	1.89	20	.94	1.71	82	51
S.D.	.86	.83		.87	.86		
Subtest 3							
Mean	1.33	1.45	9	1.18	1.86	58	33
S.D.	.79	1.03		.98	1.18		
Subtest 4							
Mean	1.09	1.73	59	1.71	2.20	28	44
S.D.	.89	1.05		1.00	1.01		

Sex Differences

Hypothesis 2 on page 13 states:

There is no significant difference between the mean performance of boys and girls on the Poetry Test.

Table XXXVIII (page 202) shows that Sex Main Effects make no significant contribution over and above grade level to the total variance of the total scores on the Poetry Test. Null hypothesis 2 is therefore not rejected. As was anticipated, boys and girls at the high school level appear to have approximately equal skill in responding to poetic language.

Reading Level Differences

Hypothesis 3 on page 13 states:

There is no significant difference between the mean performance of high Reading Test scorers and low Reading Test scorers on the Poetry Test.

In order to test this hypothesis, Grade X students and Grade XII students were assigned to top, middle, and low groups of approximately equal numbers according to Reading Test total scores. Table XXXIX (page 203) presents the results of the two-way analysis of variance on Factor A, Reading Level, and Factor B, Grade Level, for Poetry Test scores on both Test Forms.

This table shows that for Poetry Test totals the differences in responses due to differences in Grade Level are highly significant ($p < 0.001$). Differences in total Poetry Test response due to Reading Level do not attain significance at the 0.05 level for Form A but are

significant at the .02 level for Form B. Null hypothesis 3, is therefore rejected for Form B but not rejected for Form A.

The results indicate that, in line with expectations, there is a much closer relation between reading ability, as measured by the Reading Test, and the ability to respond to poetic language, as measured by the Poetry Test, than between sex and the ability to respond to poetic language.

Contrary to expectations, but reflecting the relatively low intercorrelations between Reading and Poetry Test totals reported earlier, grade level bears a far closer relation to Poetry Test total scores than do Reading Test scores. Several possible inferences can be made from the results presented in Table XXXIX. One is that general reading skill, as measured by a typical standardized reading test, while bearing some relationship to the more specialized skill involved in responding to poetic language, is sufficiently different from poetry reading skill to warrant a specialized teaching approach to poetic language. It does not appear reasonable, on the basis of the interpretation of the data relating to the present investigation, to assume that a student with good general reading ability can deal effectively with poetic language. On the contrary, it would appear that students need to be taught in a systematic manner how to read poetry.

A second inference which can be made on the basis of the above data is that the Poetry Test is somewhat unreliable. There is considerable evidence within the

present study, relating primarily to the lack of statistical equivalence of the Test Forms, to suggest some lack of reliability in the Poetry Test. Perhaps if the experimental test were more reliable, there might be a closer correspondence between its results and results on a reading test of known reliability.

A third possible inference is that the low correspondence between Reading Test scores and Poetry Test scores is due in part to the fact that the Poetry Test stresses response to individual words and expressions rather than comprehension of a whole poem. If the operation of poetic language is regarded as essentially the same as the operation of "everyday" language, it might be argued that the Poetry Test is somewhat lacking in validity, and hence will not produce responses corresponding closely with responses on a valid Reading Test. It is just this assumption which Chapter II attempted to refute, because the present study is largely based on the opposite premise, i.e., that there are fundamental differences in the operation of "poetic" and "discursive" language and that a failure to investigate and understand these differences is a major obstacle in the teaching of poetry. If these last assumptions are valid, and the Poetry Test is valid, then the relatively low correlations between Poetry Test scores and Reading Test scores reported in these results are not unexpected.

A final possible inference arising from the data presented in Table XXXIX, and to some degree from all the

data relating to the Reading Test, is that the Reading Test itself, although reliable, is not sufficiently comprehensive in measuring the various kinds of language meaning to be considered completely valid, and in fact measures only part of the total spectrum of reading ability. Particularly in the Vocabulary section, it appears to this investigator that important aspects of word meaning, chiefly relating to connotation, are not adequately measured. In determining to what extent poetic language is an extension of, rather than different in kind from, ordinary language, it is first necessary both to have an adequate conception of the nature of "ordinary language" and to have valid and comprehensive instruments for measuring response to ordinary language.

Tables LXV to LXX in Appendix D present a detailed analysis of variance results as they relate to the various subtests of the Poetry Test and the Reading Test. Although these data are generally difficult to interpret because of the lack of statistical equivalence in the test forms, two generalizations seem warranted.

First, scores in the Vocabulary subtest of the Reading Test appear to be a better predictor of success on the Poetry Test in general than are scores on the Comprehension subtest, particularly as they relate to Subtest Four (metre and rhythm). This is not unexpected in view of the fact that the Poetry Test emphasizes sensitivity to the various meanings of words and expressions more than a global understanding of "what the poem says". Thus it seems natural

that the Poetry Test would have a closer relationship to the Vocabulary rather than to the Comprehension part of the Reading Test. It is possible, of course, that the skill measured by the Vocabulary subtest is more closely related to general intelligence than that measured by the Comprehension subtest. This would also result in a greater correspondence between Vocabulary scores and Poetry Test scores than between Comprehension scores and Poetry scores.

Second, there is a tendency for Reading Test scores to be related more closely to scores on Poetry Subtests One and Four than to scores on Poetry Subtests Two (Suggestion) and Three (Sound). As Tables LXV to LXX show, in the six two-way analyses of variance involving Reading Test totals and subtotals, Reading Level (including Vocabulary Level and Comprehension Level) contributes significantly to the total variance three times for Subtest One (Thought), once for Subtest Two, not at all for Subtest Three, and three times for Subtest Four (Rhythm). Although the results do not attain statistical significance, Reading Level also makes an important contribution to the total variance on one other occasion for Subtest Four.

That success on the standardized reading test is generally a better predictor of success on Subtests One and Four than on Subtests Two and Three might perhaps be explained by postulating that a common skill, perhaps akin to analytical skill or the ability to make inferences, lies behind both Reading Test success and success on Poetry Subtests One

and Four. On the other hand, Subtests Two and Three seem to tap skills which demand a more comprehensive conception of cognition than that set forth by Bloom and his associates--skills which include ability to respond to sensory stimuli at different levels of consciousness and which are related to both the intensity of a response and the way it is organized with other similar responses for integration into the individual's total mental organization.

Arguments for the need for more effective instrumentation for exploring this latter type of mental activity (the type with which Subtest Three (Sound) is most closely associated) appear to be supported by the data presented in Tables LXV to LXX in Appendix F. Here again is an area requiring further investigation.

In examining Tables LXV and LXVI in Appendix F, which present Scheffe's Multiple Comparisons of Reading Level Effects,¹ it might be noted that these comparisons show significant differences in performance for the three reading levels (1, 2, and 3 representing the top, middle, and bottom thirds, by Reading Test totals) on Poetry Test Subtests One (Thought) and Four (Rhythm) but not on Poetry Subtests Two (Suggestion) and Three (Sound). This supports previous remarks about the various relationships between Reading Test scores and Poetry Subtest scores.

¹H. Scheffe, The Analysis of Variance (New York: Wiley, 1965), p. 112.

For performance on Subtests One and Four of Form A there is a significant difference between students in the top group (according to Reading scores) and the two other groups. For performance on the same subtest of Form B, the significant difference is between the bottom Reading group and the other two groups. A possible explanation for this result may be that Subtest One and Four items in Form A are more difficult than the corresponding items in Form B and thus more apt to distinguish between generally high-ability students and those of medium or low ability. This explanation is supported by item analysis data (see Table XI, page 106), which show Form B means for Subtests One and Four to be higher than the corresponding means for Form A.

Equivalence of Form A and B Groups

Table XLI presents a comparison of the three Reading Test scores achieved by students taking each of the Poetry Test Forms. This data is a useful check on the randomization procedure used in administering the two Test Forms. The similarity of all the pairs of figures, particularly the mean Reading totals for the total groups taking the Poetry Test Forms (73.7 for Form A as compared with 73.0 for Form B) indicates a high degree of similarity in verbal ability and reading comprehension between the A group and the B group. This similarity strongly suggests that any disparity between performance in Form A and performance on Form B is due to differences in the test forms

TABLE XLI

Reading Test Means and Standard Deviations:
 Comparison of Performance by Students
 Taking Form A and Form B of the Poetry Test

Reading Test Data	Total Group		Grade X		Grade XII	
	Form A	Form B	Form A	Form B	Form A	Form B
Vocabulary						
Mean	35.5	35.4	28.4	30.0	43.8	41.1
S. D.	14.8	12.0	11.7	7.9	13.8	13.1
Comprehension						
Mean	38.2	37.7	33.5	34.9	43.7	40.6
S. D.	12.5	10.2	13.0	9.4	9.2	10.2
Total						
Mean	73.7	73.0	61.9	65.0	87.4	81.7
S. D.	25.2	19.6	23.3	14.2	19.9	21.0

themselves rather than to differences between the two groups of subjects.

Summary of Group Performances

In an attempt to obtain further information about the skill or skills measured by the Poetry Test and its subtests and to arrive at a clearer conception of the construct under consideration, relationships between the experimental tests results and grade level, reading ability and sex were explored, primarily by means of the two-way analysis of variance procedure. The results indicate that the relation between sex and the skills measured by the Poetry Test is not significant statistically. On the other hand, the relationship between grade level and the skills measured by the poetry test is highly significant. The relationship between reading level and Poetry Test performance, as indicated by scores on the Nelson-Denny Reading Test, was chequered--sometimes highly significant, sometimes apparently important, sometimes non-existent.

Of these relationships, the one concerning reading level is perhaps the least predictable. However, it seems to be of most import in determining the nature of the construct under investigation, and is the most productive source of inferences possibly leading to further investigation.

FACTORS IN RESPONSE TO POETIC LANGUAGE

Introduction

In attempting to gain additional insight into the nature of student response to poetic language, as measured by the experimental test, the raw data were submitted to a principal components analysis¹ and a subsequent orthogonal rotation to the varimax criterion.²

The basic function of factor analysis, as noted by Eysenck, is to resolve "a set of descriptive variables into a smaller number of categories" by simplifying a correlation matrix by the use of a mathematical technique.³ "Under certain circumstances," Eysenck says, "factors may be regarded as hypothetical causal influences underlying and determining the observed relationships between a set of variables. It is only when regarded in this light that they have interest and significance for psychology."⁴ Eysenck stresses that "cause", used in this context, "is a concept which aids in the simplification and unification of natural phenomena," and that "a scientific concept is

¹Henry H. Harman, Modern Factor Analysis (Chicago: Chicago University Press, 1967).

²H. T. Kaiser, "The Varimax Criteria for Analytic Rotations on Factor Analysis," Psychometrika 23, 1958, 187-200.

³H. J. Eysenck, The Logical Basis of Factor Analysis (London: The University of London, 1967), pp. 6-9.

⁴Ibid.

not a part of nature: it is rather a way of comprehending nature." He further notes that much of the classificatory and diagnostic work in such fields as medicine "is essentially of a factorial kind, although lacking the rigor and explicitness of factor analysis," and believes that "a more formal use of these recent mathematical developments will improve more intuitive 'clinical' types of analysis."¹

Viewed in a slightly different way, factor analysis, like other statistical procedures, is a useful means both for suggesting and testing hypotheses. In performing these functions, however, as Eysenck notes, because "it provides ab initio data relevant to the formation of . . . hypotheses, and rules out a large number of possible hypotheses which might otherwise have been entertained," factor analysis appears to be of particular usefulness in a relatively unexplored field, because of its ability to "accelerate the formation of reasonable, worthwhile hypotheses, and in discarding poor ones."²

As for the testing, as opposed to the generating, of hypotheses, Eysenck notes that "there are a number of hypotheses, particularly those concerned with structure and organization, which require factor analytic methods, and which are difficult at the moment to disprove or support by nonfactorial methods. All type--and trait--hypotheses,

¹Ibid.

²Ibid.

for example, fall into this category. . . ."¹

Applying these general comments to language teaching in general and poetry teaching in particular, it would appear that factor analytic procedures have much to offer not only in supplementing present methods of generating and testing hypotheses, but also in suggesting new methods of generating hypotheses and providing previously unavailable means of testing them.

It was contended in Chapters I and II that sound teaching procedures depend on accurate knowledge of the domain and constructs relevant to what is being taught. Accordingly, the sound teaching of poetry depends on accurate knowledge of the nature of poetic language (the domain) and of the way in which pupils respond to it (the constructs). If, like I. A. Richards, we believe that diagnosis of readers' difficulties in apprehending poetic language is essential in poetry teaching, and if we can assume that accurate diagnosis is made possible only by the nosological function, by accurate categorization not so much as it applies to the specialized techniques of poetry, but as it applies to the ways persons respond to poetic language, the importance and promise of factor analytic techniques for research relating to poetry teaching seem clear. Factor analysis, in reducing the bewildering number of descriptive variables relating to something as complex

¹Ibid.

as poetic language to a limited number of categories or dimensions which may be regarded as corresponding to "causal influences" underlying the observed relationships between a set of variables (in this case responses to individual items), needs to be explored in relation to the diagnosis of pupil difficulty in reading poetry. In other words, factor analysis promises to throw light on the disease rather than the symptoms. It is difficult to cure the disease by treating the symptoms, and it seems that this, lacking the knowledge necessary to distinguish between the two, is just what poetry teachers all too often attempt to do. The lack of this kind of knowledge is precisely what Richards attempted to point up in connection with poetry, and it is unfortunate that so little follow-up has resulted from his work.

This study, like other studies of the basic nature of language and of response to language, attempts to explore what could be considered a "new" field. Because of the attendant dearth of knowledge about the construct under investigation, i.e., response to poetic language, the factor analyses performed on the test data appeared to be perhaps less profitable, in terms of clear results, but much more promising, in indicating future directions for research and for the generation of new hypotheses, than other aspects of the study. Accordingly, only a part of the considerable volume of factor analytic data generated is reported in the present chapter.

Principal Components Analysis

Principal components analysis is used in the present study to "determine the minimum number of independent dimensions needed to account for the intercorrelations among the original set of variables."¹ The problem is to identify meaningful dimensions in a multivariate domain. As Cooley and Lohnes point out, this kind of "construct-seeking" task is "most frequently accomplished today by first conducting a principal components analysis and then using the resulting principal factors as a set of references for determining the simplest structure, or most easily interpretable set of factors, for the domain in question."²

Algebraically, the basic linear model for this analysis can be represented as:

$$Z_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} \cdot \cdot \cdot \cdot a_{jr}F_{ri} \quad \cdot \cdot \cdot \quad (4.1)$$

Where

Z_{ji} is the standard score for individual i on test j .

a_{jr} is the component (factor) loading of test j on component (factor) r .

F_{ri} is the component (factor) score for individual i for component (factor) r .

The observed variables (test items) are thus described

¹Wm. Cooley and Paul Lohnes, Multivariate Procedures for the Behavioral Sciences (New York: Wiley, 1962), pp. 158-61.

²Ibid.

linearly in terms of factors, or components. Each additional factor makes a maximum contribution to the total variance of the remaining variables.

The above model can be expressed in matrix terms as:

$$Z = AF \quad . . . (4.2)$$

where

Z = standard score matrix of order $(n \times N)$

A = a matrix of component loadings of order $(n \times r)$

(see Tables XLII, XLIII)

F = a matrix of component scores of order $(r \times N)$

and

N = number of observations

n = number of variables

r = number of components

If $Z = AF$ is postmultiplied by its transpose and then divided by N , the correlation matrix R is formed:

$$\frac{Z Z'}{N} = \frac{A F F' A'}{N} \quad . . . (4.3)$$

$\frac{F F'}{N} = I$, the identity matrix, if the component scores are scaled with zero mean and unit variance and are orthogonal.

Under these conditions, expression 4.3 reduces to $R = A A'$.

The matrix A is then determined by a principal components analysis of R .

The Varimax Rotation

In multiple factor analysis, a procedure such as the

principal components analysis is frequently followed by a rotational scheme designed to "convert the obtained principal factor pattern to a pattern of simple structure." One of the most commonly used of these rotation schemes, the Varimax solution,¹ was selected for the present study because it has the advantage of "cleaning up" factors rather than variables, i.e., it tends to yield high loadings on each factor for a small number of variables and very low loadings on the remaining variables. This loading pattern tends to simplify identification and interpretation of the factors.

Factor Analysis Results

Strictly speaking, the experimental test designed for the present study fails to meet optimal criteria for productive factor analyses.² The sample is of insufficient size, there are too few items in each test form, test reliability does not appear to be sufficiently high, and there is a lack of clear cut hypotheses. These deficiencies, however, are to a large extent due to the nature of the area tested and to the basic form of the test. While the usefulness of factor analysis for the present study may be limited, thus prohibiting a clear delineation and description of factors in response to poetic language, certain

¹Kaiser, op. cit., pp. 187-200.

²Guilford, Psychometric Methods (New York: McGraw-Hill, 1954), p. 531.

tentative interpretations seem warranted.

An extensive examination was made of rotated factor loadings of each of the items in each test form, proceeding by the usual method of noting the items with substantial loadings on a given factor in contrast with other items with low loadings in the attempt to discover common features of the cluster of items not shared by another group. Analyses involving various rotational transformations of the principal components solution based on both Pearson and tetrachoric item intercorrelations were performed, first for four factors, or dimensions, and then for three factors. Separate analyses involving all the above mentioned variations were performed, first on total test data for each form, then for data provided, successively, by each of the grade groups, by the top-half scores and bottom-half scores, and by various combinations of subtest scores.

Interpretation of the data generated by all these procedural variants appeared most consistent for the three-factor analysis of the Pearson product-moment item intercorrelations for the whole-group data.

Aside from considerations of interpretability, the decision to retain three factors was influenced by the fact that the addition of a fourth factor did not appreciably increase the per cent of trace.

Tables XLII and XLIII present the respective results produced by this procedure for Form A and B items.

Table XLII presents eigenvalues, or latent roots,

TABLE XLII

Eigenvalues, Per Cent Trace, and Varimax Rotated
Factor Coefficients for the Total Form A
Pearson Item Intercorrelations

First Six Eigenvalues					
1.93	1.55	1.47	1.28	1.16	1.16
Per cent of Trace of First Three Factors = 30.93					
Varimax Rotated Factor Coefficients					
	Communalities	I (B)	II (A)	III (C)	
1	.55	-.28	.46	.51	
2	.36	.09	.59	-.07	
3	.37	.19	.11	.57	
4	.48	-.01	-.53	.45	
5	.29	.10	.52	.03	
6	.18	.33	-.27	.02	
7	.24	.11	.48	.06	
8	.35	.37	.07	.46	
9	.22	.47	-.02	.03	
10	.48	-.28	-.11	.63	
11	.20	.43	.06	-.12	
12	.49	.65	.12	.21	
13	.31	.55	-.01	.10	
14	.16	-.05	.36	.18	
15	.10	-.24	-.19	.06	
16	.16	.37	.19	-.10	
	4.95	1.80	1.63	1.52	

per cent of trace of the first three factors, and varimax rotated factor coefficients for Pearson correlations for total group scores on Form A items.

From this table it will be seen that items 9, 11, 12, and 13 load high on Factor I; items 6, 8, and 16 load fairly high on Factor I; and items 1, 10, and possibly 15 load low on Factor I. Items 1, 2, 5, and 7 load high on Factor II; item 14 loads fairly high on Factor II; and items 4 and 6 load low on Factor II. Items 1, 3, 4, 8, and 10 load high on Factor III; items 2, 11, and 16 load fairly low on Factor III.

Clearly, these factor groupings cut across the four test categories and subtests. There appears, for example, to be no single factor equivalent to "metrical sense", or to "sensitivity to sound or connotative values". These factors seem rather to reflect different modes, or mental strategies which students use in matching these parts of poems to their respective wholes. Factor I on Form A, thus, appears to load heavily for items (a) where this matching involves more than one type of clue-ing, (b) where the clues tend to appear in the fairly immediate context, and (c) where both a sense of the pattern of the context and a sensitivity to the flavour of the part appear about evenly necessary. Factor I appears to involve a blend of a conscious, analytical process and an unconscious, intuitive process, operating at various degrees of subtlety.

Item A-9 ("Railway Train") for example involves

careful attention partly to the suggestion, but mainly to the sound, both of "peer", the "correct" alternative, and its immediate context, particularly "pare", the word with which it rhymes. Similarly Item A-11 ("Fear no more") requires an awareness of the connotations of "reed", the part, and of the semantic requirements of the context, which calls for an opposite to "oak". Item 13 ("I did not lose my heart") requires not only awareness of metrical pattern, but a sensitivity to a subtle variation within this pattern, in this case a feminine ending, and to a subtle difference in suggestion among the alternatives.

Unlike Factor I, Factor II for Form A appears to have high coefficients for items where (a) the matching between part and whole involves only one type of clue-ing, (b) the clues tend to appear in recurrent patterns or in the larger context of the whole poem, and (c) a sense of the pattern (this could involve such things as metre, dramatic situation, recurring symbols, sustained metaphor, etc.) is the main element. Factor II appears to involve a highly conscious, analytical process, comprising several degrees of subtlety.

As an example of an item loading high on Factor II, Item A-2 appears to require an understanding of the general dramatic situation in the whole poem and of the denotative meaning of the alternatives. Here inferring the correct meaning of the whole is considerably more demanding than understanding the meaning of the part. Item A-7 (from

"Tintern Abbey") also appears to involve matching using only one type of clue, in this case a conventional blank verse pattern in the context.

Factor III appears to load heavily for items where (a) the part-whole matching process involves "stock response" associations and/or the routine applications of simple formulae, (b) the clues are taken mainly from the "feel" or "flavour" of the alternative itself, and to a very limited degree from the context (where they are taken from the context, they are taken from its obvious aspects, such as simple word repetition), and (c) a sense of the normal flow patterns of ordinary language is a main element. Factor III appears to be related to a largely unconscious, intuitive process.

In Item A-10, for example, selecting the correct alternative, contrary to the intent of the item, is simply a matter of matching "Christmas" with "Peace and Goodwill". Item A-3, intended to measure understanding of the meaning of the complete sonnet, appears rather to measure on intuitive feel for the sound and suggestion of "loss of thee", the "correct" alternative.

Table XLIII presents data for Form B items. From this table it will be seen that items 7, 9, 10 and 14 load high on Factor I; items 11, 13, and 16 load fairly high on Factor I; items 2 and 12 load fairly low on Factor I; and item 6 loads low on Factor I. Items 2, 8, and 12 load high on Factor II; items 11, 14, and 15 load fairly high on

TABLE XLIII

Eigenvalues, Per cent Trace, and Varimax Rotated
Factor Coefficients for the Total Form B
Pearson Item Intercorrelations

First Six Eigenvalues					
2.65	1.57	1.44	1.32	1.20	1.17
Per cent of Trace of First Three Factors = 35.24					
Varimax Rotated Factor Coefficients					
	Communalities	I (A)	II (B)	III (C)	
1	.34	.13	-.10	.56	
2	.49	-.07	.68	-.16	
3	.18	.12	.19	.36	
4	.25	.07	.03	.50	
5	.34	.28	.27	.43	
6	.62	-.32	-.15	.70	
7	.56	.68	-.30	.08	
8	.53	.17	.64	.32	
9	.50	.69	.13	-.07	
10	.30	.52	-.12	.13	
11	.35	.30	.31	.41	
12	.37	-.08	.60	.02	
13	.16	.29	.17	.20	
14	.35	.49	.33	.06	
15	.15	.18	.34	.06	
16	.16	.37	.11	.09	
	5.65	2.05	1.85	1.75	

Factor II; items 1 and 10 load fairly low on Factor II; and item 7 loads low on Factor II. Items 1, 4, 5, and 6 load high on Factor III; items 3, 8, and 11 load fairly high on Factor III; and items 2 and 9 load fairly low on Factor III.

These factor groupings, like those arising from Form A, cut across the four test categories. They also seem to reflect the three types of mental strategies described in connection with the Form A data. However, while Factor III appears to be essentially the same in both Forms, Factor I and II for Form A become Factors II and I for Form B. Because of this apparent reversal of order, the basic factors or constructs underlying the factor structure of both test Forms have been labeled A, B, and C. A corresponds to Factor II on Form A and Factor I on Form B. B corresponds to Factor I on Form A and Factor II on Form B. C corresponds to Factor III on both Test Forms.

Table XLIV presents a conceptual model based on the data from Tables XLII and XLIII, and confirmed by the data presented in Tables LXXI and LXXII in Appendix F. Table XLV lists items from each Test Form having high or low loadings on any of the three factors.

To give a clearer indication of the characteristics of each of the Factors A, B, and C and of the process which led to their general description, as well as to support the conceptual scheme presented in Table XLIV, items with high or low loadings on each of the factors are now discussed in some detail.

TABLE XLIV
Description of Factors A, B, and C

Criteria	Factor A	Factor B	Factor C
Level of Awareness	High (Analytical)	Medium (Partly Analytical)	Low (Mainly Unconscious)
Type of Clueing	Uni-Level (use of one type of clue in context)	Multi-Level (use of various types of clues in context)	Stock-Association (application of simple formulas)
Use of Context	Use of Larger Context of Recurrent Patterns	Use of Proximate Context--Reliance on both context and "feel" of the part	Minimal use of context or use of only very obvious aspects of context. Major reliance on "feel" of the part itself.
Difficulty	Medium Difficulty (one screen)	Most Difficult (more than one screen)	Least Difficult (little screening)
Major Stress	Pattern, Form	Both pattern and instinctive uses of language	Habitual Uses of Everyday Language

TABLE XLV

Items from both Test Forms with Varimax Rotated
Factor Coefficients of More than .30 and Less
than -.25 for Pearson Item Intercorrelations

Item Source	Factor A		Factor B		Factor C	
	High	Low	High	Low	High	Low
Form A Item Numbers	1*,2, 4,6 5,7, 14		6,8*, 1,10 9,11, 12, 13,16		1 ^x ,3, 2 ⁺ , 4,8 ^x , 11 ⁺ , 10 16 ⁺	
Form B Item Numbers	7,9, 6 10, 11 [≠] , 14 ^x , 16		2,8*, 7 11*, 12, 14*, 15		1,3, 2 ⁺ , 4,5, 9 ⁺ , 6,8*, 11	

Note:

- x = Items with fairly high loadings on other factors
- * = Items loading higher on one other factor
- ≠ = Items loading higher on two other factors
- + = Items loading low, but higher than -.25

Items with significant Factor A loadings. Of the 32 items in the total Poetry Test (Form A and B), 11 had Varimax rotated factor coefficients of 0.30 or more for Factor A. This compares with 13 items for Factor B and 12 items for Factor C. As can be seen by examining the unrotated factor matrices for Forms A and B (Tables LXXI and LXXII in Appendix F) this similarity in the number of items with high loading reflects the tendency of the Varimax rotation to redistribute the loadings more evenly on all factors, so that it would be misleading to conclude from these Varimax figures that each of the three factors accounts for a fairly equal per cent of trace.

Of the total of 11 items with high coefficients on Factor A, one (B-11) loaded equally high or higher on two of the other factors, making interpretation difficult. Of the remaining 10 items, two (A-5 and A-7) involve a metrical pattern; two (B-7 and B-14) involve both metrical and stanzaic patterns; one (A-1) involves a conventional rhyming pattern; two involve the total dramatic or plot pattern of a poem (A-2 and B-13), two appear to involve either simple word recurrence (B-9) or recurrence of words with similar meaning (B-16); and one (A-14) appears to involve an alliterative patterning. The significant common element among these items seems to be the relative importance of a particular contextual pattern in matching the part to the whole. This seems to apply whether the difficulty level of an item is high, as in A-14 (0.07) or low, as in

A-1 (0.76).

Items with low coefficients on Factor A (A-4, A-6, B-6) confirm this "pattern" hypothesis. For these items the matching process appears to be governed by the very strong suggestive and sound values of the part as opposed to the pattern of the whole.

Items with significant Factor B loadings. Of the 13 items with high coefficients on Factor B, two (B-11 and B-14) also show high loadings on Factor A; three (A-8, B-8, and B-11) show high loadings on Factor C; and one (B-11) shows high loadings on both of the other factors. Factor B appears more complex than Factors A and C, involving a blend of pattern recognition and sensitivity to the flavour of individual words and expressions.

Of these 13 items, three (A-13, A-16, and B-14) were intended to measure response to a metrical pattern; two (A-12 and B-8) appear to relate to "tone"; two (A-9 and B-15) relate to half-rhyme, or slant-rhyme; three seem to relate to compact verbal patterns, involving metaphor (B-14) and analogy (A-11 and B-2); and three involve the sound and connotative meanings of words and expressions as these meanings relate to the more-or-less immediate context (A-6, A-8, and B-12). It would at first seem likely that the "metre" items (A-13, A-16 and B-14) would load high on Factor A rather than Factor B, but a close examination of these three items reveals that in two cases (A-13 and A-16)

the metrical distinction between the two likeliest alternatives is extremely subtle and that in the remaining case (B-14) the metrical pattern seems too involved for the student to cope with.

Thus, although these last three items purport to measure sensitivity to the metrical use of language, they may in fact measure sound-connotation values. The same transfer seems to operate with respect to items A-9 and B-15, which were intended to measure awareness of half-rhyme. Because students were virtually unaware of the operation of half-rhyme, selection again appeared to be made on the basis of the sound-connotative values of the various alternatives.

In cases such as the ones discussed in the preceding paragraphs, where what is being measured is of a highly complex nature, factor analytic techniques promise to be an invaluable complement to intuitive judgment in indicating, and hence diminishing, the discrepancy between what an item purports to measure and what it actually measures in terms of underlying psychological constructs. It appears from the results emerging from the present study that face validity, whether or not supported by knowledge of subject matter, is not sufficient by itself in this kind of testing situation.

Items with low loadings on Factor B (A-1, A-10, and B-7) also serve to support the description given above and presented in Table XLIV. These items clearly fail to require any significant degree either of analysis or

word-sensitivity. Rather, they appear to involve the routine application of a formula (A-1), a simple stock response (A-10), or the recognition of a very simple pattern (B-7).

Items with significant Factor C loadings. Of the 12 items loading high on Factor C, two (A-1 and A-8) appear to involve the routine application of a simple formula. For A-1 this formula has to do with rhyme and for A-8, with alliteration. Item A-10 clearly involves the routine association of "Peace and Goodwill" with "Christmas", and item B-3 possibly involves the same kind of routine association, this time of "cars" with "lights". Both items were designed to measure ability to make an inference from the total situation presented by a poem, but this intent was apparently frustrated by the "stock response" phenomenon which has been noted by Richards and others.

The remaining eight items appear to involve the selection of an alternative with little or no concern for the context or its contribution to the total poetic effect. Rather, the selection in these items seems governed primarily either by the attractiveness (vigour, smoothness, etc.) of the alternative in itself or by the ease with which the alternative fits in with the flow of everyday speech. The former effect is seen in items A-4, B-5, B-6, B-8, and B-11; the latter in items B-1, B-4, and perhaps A-3. It also seems likely that both these influences are at work in

some items, such as A-3 and B-1.

Because of insufficient data these remarks, particularly as they concern items with high Factor C loadings, are highly tentative and speculative, and subject to further more detailed investigation.

Items loading relatively low on Factor C (A-2, A-11, A-16, B-2, B-9) all appear to involve a selection involving an analysis of context and a high degree of awareness of various kinds of patterns in the context. That these items all load high on one of the other two factors would seem to support the scheme set forth in Table XLIV. Item A-2, for instance, which loads low on Factor C loads high on Factor A. This item ("The Fiddler of Dooney") seems to be concerned with a fairly high degree of awareness of the total meaning of the poem (Factor A) and not at all concerned with any kind of stock response or unconscious awareness of sound or suggestiveness in the alternatives (Factor C). Item B-2, which also loads low on Factor C, loads high on Factor B. This item ("Others, I am not the first") is concerned both with the completion of an analogy (ice is to fire as _____ is to desire) and with the association of "ice" with "fear" rather than "guilt" (Factor B). No application of a simple formula (Factor C) will serve in this case.

It should be pointed out that all the items, and not merely those examined in the following section, load to some degree on more than one factor. This would suggest

that the item interpretations are complex, and that each item involves two or three (or more) complex psychological continua in its interpretation. In the light of the complexity of poetic language the irregular pattern of factor loadings for the present items is not unexpected.

Items with high loadings on all factors. An examination of items B-5 and B-11, which load high on all three factors, seems to lend further support to the contention that poetry interpretation is complex in nature. In item B-5, the selection of the "correct" alternative could have reasonably been based on a large number of different types of clues. The intent of the item was to measure students' ability to match the sound-connotative value of "gleaned my teeming" with similar values running throughout the whole poem. But the actual selection of the response could have been based on the denotative connection between "gleaned" and "grain", or simply on the attractiveness (smoothness, richness, musical quality) of the phrase itself. The same considerations apply to item B-11, where "coil", the "right" answer, could have been selected because of the sinister mood of the whole situation developed in the poem, or simply because snakes hiss and coil and strike at people's legs. Thus for both items selections are likely to be made based on very different uses of context and at very different points on an analytical-intuitive continuum. Although of limited usefulness in delineating and describing

clear-cut factors, such items are interesting as traps, or catch-alls, for various mental strategies, and perhaps serve as a useful reminder that there is no necessarily "right" way of getting meaning from poetic language.

Interpretation of the Factors in the Light of Poetic Theory

Introductory comment. C. S. Lewis, in distinguishing between "literary" and "unliterary" readers,¹ speaks of "receiving" the stimulus produced by a work of art as opposed to merely "using" it. As examples of "using" he cites extracting a tune or a vague emotion from a symphony, a sentiment from a painting, a plot from a novel. This "using", he says, is not a bad thing in itself; it simply fails to ask of a work of art the main thing, which could be called aesthetic experience, it has to offer. In order to participate in this experience the viewer, or listener, or reader, must "receive". In the case of a poem, he must allow the words to do their work on him. He must approach poetic language in a state of high awareness but also in a state of passivity. One of the central assumptions underlying this investigation is that a basic obstacle to adequate response to poetic language is a faulty set of expectations on the part of the reader regarding the function of words as they occur in a poem. The poet, by the nature of his art, must deal in words that are most commonly used for their power of immediate association. Normally words

¹C. S. Lewis, An Experiment in Criticism (Cambridge University Press, 1961), Chapter I, pp. 139-40.

are not savoured or seen as a source of new meanings or new experience. Their role in the business of daily survival tends to forbid this.

Not so common, but still much more common than the poetic use of words, is their "scientific" use. Here precision of meaning and a high degree of conscious analysis are required. Where daily living tends to train people in the "stock-response" use of words, formal education tends to train them in this "scientific", or analytical, or logical use.

The third use of words could be termed "poetic" or "literary". Because it is not so obviously necessary for survival, or because it often runs counter to the other uses, or because in incorporating and surpassing the other uses it demands more of readers and listeners, it tends to be neglected.

The importance of this neglect must be determined by a society as a whole, but its exact nature and cure (if this is what is desired) is the responsibility of critics, teachers, and researchers. In any case, no remedy for this situation is possible without much additional information regarding the nature of response to language in general and to poetic, or literary, language in particular.

The nature of the three factors. In undertaking an investigation of the nature of response to poetic language, this study, because it is largely exploratory, is concerned

with generating hypotheses for subsequent investigation rather than with testing hypotheses. The factor analytic procedures described in the preceding section produced three factors which appear to be amenable to a conceptual scheme which has a fairly high degree of face validity and internal consistency. This scheme confirms reasonably well the ideas concerning the nature of poetry presented in Chapter II and appears successful in suggesting certain possibly useful hypotheses.

In terms of the three broad uses of language, the "everyday", the "scientific", and the "poetic", there appears to be a strong connection between Factor A and the "scientific" use of words, between Factor B and the "poetic" use of words, and between Factor C and the "everyday" use of words. As is shown in Table XLIV (page 229), Factor A appears closely associated with a highly analytical kind of response appropriate to the "scientific" uses of language; Factor B with the multi-level type of response appropriate to "poetic" uses of language; and Factor C with the mainly routine, conscious type of response appropriate to the "everyday" uses of language.

On the basis of the conceptual pattern suggested by these factors it should prove possible to devise measuring instruments, perhaps similar to the one used in the study, in which individual items are systematically constructed to load high on the respective factors. Subsequent analysis might suggest additional component factors and generate

further hypotheses.

In the present study, Factors A and B, corresponding to "scientific" and "poetic" uses of language, involve items which are considerably more difficult than Factor C (everyday language) items. Also, there is a greater performance gain between Grade X and Grade XII for items loading high on Factor A, and a much greater performance gain for items loading high on Factor B than for items loading high on Factor C. These findings, presented in Table XLVI, tend to confirm the conceptual scheme outlined earlier. It might be noted that by far the highest Grade X-Grade XII gain is for items loading high on Factor B, which has been tentatively identified as corresponding more closely than the other factors to the "poetic" uses of language.

Summary of Findings Relating to Factor Analysis

As a result of the factor analysis, certain main findings emerged:

1. Three reasonably distinct factors in response to poetic language as measured by the experimental Poetry Test were identified and tentatively described. That these three factors account for only 31 per cent of trace for Form A and 35 per cent of trace for Form B strongly indicates the presence of further, somewhat less pronounced factors within the construct under investigation.

2. The factor loading pattern for the various items

TABLE XLVI
Difficulty Level Change from Grade X to Grade XII
for Items with High Factor Loadings

Factor A			Factor B			Factor C		
Item No.	Difficulty X	Per cent XII Increase	Item No.	Difficulty X	Per cent XII Increase	Item No.	Difficulty X	Per cent XII Increase
A1	.74	.80	A6	.29	.30	A1	.74	.80
A2	.29	.48	A8	.45	.48	A3	.25	.72
A5	.33	.37	A9	.29	.48	A4	.49	.46
A7	.29	.41	A11	.41	.26	A8	.45	.48
A14	.06	.09	A12	.16	.48	A10	.63	.59
B7	.51	.70	A13	.16	.30	B1	.65	.66
B9	.27	.47	A16	.28	.58	B3	.61	.75
B10	.49	.62	B1	.65	.66	B4	.63	.64
B11	.27	.45	B8	.25	.51	B5	.16	.38
B14	.08	.17	B11	.27	.45	B6	.45	.55
B16	.39	.44	B12	.08	.28	B8	.25	.51
			B14	.08	.17	B11	.27	.45
			B15	.26	.46			
Means	.34	.45		.28	.42		.46	.58
		32			50			26

is irregular and difficult to interpret. A few items have high loadings on all three factors, and several items have high loadings on two factors. Most of the 32 items show an appreciable degree of load on more than one factor. This complexity appears to reflect the complexity of the skill or set of skills measured by each item and is not unexpected in view of the intricacy of poetic language.

3. The irregularity of the factor-loading pattern is also reflected in the outcome that the three factor groupings largely fail to parallel the four test categories. This, more than anything else, makes accurate description of the three identified factors difficult, and suggests a serious lack of correspondence between the intent of the various items and the process of responding to them.

4. The skills associated with the three identified factors were tentatively characterized as follows:

Factor A: Highly analytical, high level of awareness; use of single-level clue-ing, use of larger context or recurrent patterns; major stress on pattern and form, and emphasis on the "whole" rather than the "part".

Factor B: Partly analytical, partly intuitive; fluctuating levels of awareness; use of proximate context; generally difficult; equal reliance on "whole" and "part".

Factor C: Instinctive; low, mainly unconscious level of awareness; use of stock association and the

application of simple formulas; minimal use of context, with major reliance on the "feel" of the "part"; generally easy; associated with the habitual uses of "everyday" language.

5. The three factors described above appear to parallel the "scientific use" of language (Factor A), the "poetic use" of language (Factor B), and the "everyday use" of language (Factor C).

6. The findings resulting from the factor analysis confirm the commonly held critical view (see Chapter II) that in each component of poetic language, even at the single-word level, several kinds of meaning are operating simultaneously.

CHAPTER SUMMARY

The findings relating to (1) the performance of the experimental poetry test, (2) the performance by certain categories of individuals on the test, and (3) factors in response to poetic language, were presented in this chapter. A brief review of these three groups of findings, summarized previously on pages 195, 214, and 240, is presented as follows:

1. In view of the complexity of poetic language, the Poetry Test appears to be adequately reliable for investigative purposes.

2. Examination of detailed item data reveals that the testing procedure used in the study is effective in

providing a wide range of reasonably exact information, at varying levels of sophistication, about the component skills in responding to poetic language.

3. A possible weakness in the present testing procedure is that the interdependence of the elements operating in poetic language make it extremely difficult to construct items which reliably measure discrete skills. This is particularly true for the "Sound" and "Suggestion" categories.

4. The generally low degree of correspondence among performances on the four subtests (the mean correlation coefficient is less than 0.2) indicates that the categories represent skills which, though interrelated, are decidedly discrete.

5. The low correlation between total Poetry Test performance and Reading Test performance suggests that skills involved in responding to poetic language are substantially different from those involved in responding to discursive language at least, as measured by standardized reading tests. Also, the relatively greater between-grade gains for the seemingly more reliable of the Poetry Test forms suggest that high school students' skill in responding to poetic language is increasing at a greater rate than their skill in responding to discursive language, perhaps mainly because the latter skill is already relatively well developed.

6. A breakdown of Poetry Test scores by grade, by

sex and by Reading Test performance reveals that grade level is by far the most important variable, considerably more closely related to Poetry Test performance than Reading Level, and far more influential than Sex Difference, the effect of which is virtually non-existent.

7. Generally the test is too difficult for the average student in the sample tested. Accordingly, it appears more suitable for Grade XII than for Grade X students.

8. Three factors in response to poetic language, as indicated by Poetry Test results, were identified and tentatively described. That these factors failed to parallel the existing test categories, and that the factor loading pattern is complex, with many items having substantial loadings on more than one factor, suggests that each item measures a complex of skills.

9. The three factors which were identified were tentatively labeled Analytical (Factor A), Analytical-Intuitive (Factor B), and Intuitive (Factor C). These factors appear to correspond with the "scientific", "poetic", and "everyday" uses of language.

The following chapter presents the results and discussion arising from the follow-up interviews which accompanied the administration of the Poetry Test.

CHAPTER V

INTERVIEW RESULTS

INTRODUCTION

The main purposes of this study are first, to investigate student response to poetic language and second, to examine and develop means for examining this response. In accordance with these purposes an experimental poetry test was devised and administered to high school students and test results were analyzed in an attempt to discover possible patterns and relationships concerning response to poetic language. As a result of these analyses it was possible to suggest some tentative generalizations relating to why certain alternatives were chosen. Testable hypotheses for further investigation could be formulated from these generalizations. This will be discussed in the final chapter.

The purpose of the follow-up interviews is to supplement the knowledge gained from the analyses of Poetry Test results presented in Chapter IV. Shortly after the administration of the test students were questioned about their own responses to the various test items. It was hoped to gain, through respondents' direct perceptions of their own acts of selection, valuable additional knowledge

about the students' thought processes while reading a poem. At the same time the interviews would serve to provide additional information as to the validity and reliability of the Poetry Test itself.

A description of the manner in which the interviews were conducted was presented in Chapter III. Briefly, the interviewing procedure consisted of a five-minute period for establishing an appropriate climate, a forty-minute period in which students were asked to explain the reasoning behind the selections in the various items, and a final five-minute period for any free response not already offered. While the questions and questioning sequence in the forty-minute "body" of the interview were structured and similar for each respondent, free response was encouraged at all times. It was hoped that the resulting balance between structured questions and free response would produce an optimal condition for the gathering of information.

As was described in Chapter III, the interview sample was selected so as to represent students achieving low, medium, and high scores on the Poetry Test. The sample of ten students included two Grade XII students with very high scores (98th percentile); one Grade XII and three Grade X students with high scores (above the 80th percentile); one Grade XII student with a score at the 50th percentile; one Grade XII and two Grade X students with scores between the 20th and 40th percentiles. These percentiles were calculated within each Grade group.

Typescripts of two complete interviews, chosen from among the ten because of the ability of these respondents to express themselves, are presented in Part C of the Appendix.

INTERVIEW RESULTS AND THE POETRY TEST

Test Validity

The results of the ten interviews conducted with students selected to represent a range from below average test scorers to very high scores tend to confirm the general validity of the experimental Poetry Test. As Table XLVII indicates, the majority of the items on both Test Forms appeared to measure what they were intended to measure. Essentially, the interviews attempted to discover to what extent students could explain their selections within the various items and to investigate the reasoning behind each selection. Because the Poetry Test required students not only to make a selection within a given dimension, such as "tone", but also to determine which dimension was being measured before making this selection, it would seem reasonable that a high correlation between the selection of "correct" alternatives and the selection of these "correct" answers for the "correct" reasons would indicate a high degree of content validity.

The data presented in Table XLVII were tabulated as follows. During the interviews three notations, one noting whether or not the original selection was accurately recalled, the second noting the "correctness" of this

TABLE XLVII

Recall, Correctness, and Dimension of Test Responses as Indicated
by the Follow-up Interview: Detailed Comparisons

Item Number	STUDENT BY NAME AND GRADE											
	from Form B											
	Katherine (XII)			Ann (XII)			Lance (XII)			Carol (X)		
	R	C	D*	R	C	D	R	C	D	R	C	D
1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1
6	0	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1
Total	15	14	15	12	14	14	12	14	8	14	10	15
										6	8	8

* R = original response correctly recalled
C = original response "correct"
D = original dimension used "correct"

† = the borderline cases were designated "1/2"

TABLE XLVII (Continued)

Item Number	STUDENT BY NAME AND GRADE															
	from Form B				from Form A											
	Greg (XII)				Gordon (X)				Veronica (X)				Elaine (X)			
	R	C	D		R	C	D		R	C	D		R	C	D	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
5	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Total	13½	12	12½		13	5	7½		13½	9	11½		10	5	4½	

original selection, and the third noting the "correctness" of the dimension used in making the response, using "1's" and "0's", were made for each item for each interviewee. Table XLVII is a compilation of these results, containing 480 pieces of information, 48 for each interviewee. From the data presented in Table XLVII it was estimated that the correlation between "correct" responses and "correct" reasoning leading to these responses, exceeded 0.90. Put in a slightly different way, 79.5 out of the 92 correct answers reported in Table XLVII, or 87 per cent, were made by students using the "correct" dimension, while only 12.5 correct answers, or 13 per cent, were made by students not using the "correct" dimension. On the other hand, of the 68 "incorrect" answers reported in Table XLVII only 28, or 41 per cent, were made by students using the "correct" dimension.

An examination of the data presented in Table XLVII further shows that a total of 107 responses were made by employing the "correct" dimension. Of these 87, or 75 per cent, produced "correct" answers while only 27, or 25 per cent, produced "incorrect" answers. However of the 53 responses not made by employing the "correct" dimension only 13, or 25 per cent, produced "correct" answers, while 40, or 75 per cent, produced incorrect answers.

The relationship between correctness of response and correctness of dimension was calculated in the same manner for the four subtests and summarized in Table XLVIII. It

TABLE XLVIII

Comparison of Correctness of Response with
Correctness of Dimension for Poetry
Test Totals and Subtotals

Test Answers	Percent of Response Based on "Correct" Dimension				
	Total	ST1	ST2	ST3	ST4
Correct	87	90	80	62	80
Incorrect	40	/	/	/	/

/ these percentages were not computed.

TABLE XLIX

Comparison of Correctness of Response with
Correctness of Dimension for Items Loading
High on Factors A, B, and C

Dimension Used	Percent of "Correct" Answers		
	Factor A Items	Factor B Items	Factor C Items
Correct	84	79	68
Incorrect	14	21	31

was reasoned that Subtests One and Four, meant to measure responses which are more analytic and less intuitive than responses measured by Subtests Two and Three, would show a correspondingly greater ability on the part of students interviewed to give "correct" reasons for arriving at "correct" answers. Of the total number of "correct" answers to Subtest One ("Thought") items reported in Table XLVII, 90 per cent were made for the "correct" reasons. The corresponding figures for Subtests Two, ("Suggestion"), Three ("Sound") and Four ("Rhythm") are 80 per cent, 62 per cent, and 80 per cent, respectively. The fact that the Subtest One figure is noticeably higher than the Subtest Three figure appears to support the notion that "meaning" items involve a higher degree of conscious thinking than "sound" items and thus might be expected to result in a greater tendency to give "right" reasons for "right" choices. This particular comparison of the subtests indirectly lends support to the validity of the Poetry Test.

Another approach which appears potentially useful involves comparing the relationship between "correctness" of answers and "correctness" of the dimension within which selections were made (see Table XLVIII) with the factor loading results presented in Chapter IV. This comparison is intended to illuminate and perhaps substantiate the interpretational scheme for factor loadings. Table XLIX presents this comparison in summary form. For the three items from both Test Forms which loaded high on "Factor A",

84 per cent of the "correct" answers coincided with the "correct" dimension. For the four "Factor B" items the corresponding figure is 79 per cent and for the three "Factor C" items it is 68 per cent. Looked at in another way, 14 per cent of the Factor A "correct" responses, 21 per cent of the Factor B "correct" responses, and 32 per cent of the Factor C "correct" responses were based on the "wrong" dimension.

These differences are considerable and appear to support the tentative conceptual scheme advanced in Chapter IV to explain results of the factor analysis. If Factor A, as was suggested, involves a relatively analytic, highly conscious mental process, and Factor C involves a kind of intuitive associational response with a relatively low level of analysis, it seems reasonable to assume that interviewees would be more readily able to give "right reasons for right responses" for Factor A items than for Factor C items. This assumption was borne out by the figures presented above. Further, if Factor B, conceptually, lies between the other two factors, the 79 per cent figure for Factor B items (as opposed to 84 per cent for Factor A items and 68 per cent for Factor C items) could reasonably have been anticipated.

The interview results, therefore, appear to support the factor analysis results and interpretation presented in Chapter IV. Again the validity of the Poetry Test seems to be supported.

One further approach relating to the data presented

in Table XLVI is to attempt to determine whether the test is more valid for high scorers than for low scorers. It could be argued that purely random guessing is more prevalent in all types of objective tests for low scorers than for high scorers and that hence test validity would be higher for high scorers in all objective testing. Since, more than most, the present test is more directly concerned with investigating student response than with evaluating pupils, and uses "correct" answers as an indirect means of making inferences about response patterns, it is particularly valuable to ascertain the accuracy with which "correct-response" patterns reflect "correct approach" patterns.

It has been suggested that for high scoring students the degree of coincidence between "correct approach" and "correct response" should be much higher than for low scoring students and that in this sense, test results as a whole would be more valid for the high scorers. This hypothesis was tested by comparing the relationship between "correct" answers and "correct" dimension for the two interviewees with the highest scores and for the two with the lowest scores. Table XLX presents a summary of this comparison. For the high scoring students, 95 per cent of the "correct" answers and 75 per cent of the "wrong" answers involved the "correct" dimension, while for the low scorers, only 50 per cent of the "correct" answers and 35 per cent of the "wrong" answers involved the "correct" dimension. These figures clearly support the idea that there is a closer

relationship between "correct" selections and "correct" reasons for making these selections for high scorers than for low scorers. In terms of the purposes of the present test, therefore, individual items generally appear more valid for high scoring than for low scoring respondents.

It should be noted that in the discussion of interview results "dimension" denotes the student's strategy of approach to an item, and corresponds to one of the four test categories. For example, if a student approaches a "rhythm" item on the basis of rhythm, or metre, he is using the correct "dimension". As used here, "dimension" is not to be confused with "factor analytic dimension".

However, because of the small number of students interviewed, the foregoing results must be viewed with caution. These results can only be taken as indicating a probable trend, and are obviously far from conclusive.

TABLE XLX

Comparison of Correctness of Response
and Correctness of Dimension
for High and Low Scorers

Test Answers	Percent Using "Correct" Dimension	
	High Scorers	Low Scorers
"Correct"	95	50
"Incorrect"	75	35

Test Reliability

In addition to providing additional information about the validity of the experimental test, the data presented in Table XLVII provide information about its reliability. Because of the lack of homogeneity, the limited number of items in each test Form, and because the two Forms, which were meant to be equivalent, were discovered to be only partially parallel, standard procedures for determining reliability may be applied to the Poetry Test to only a limited degree. The interviews, a part of which consisted essentially of retesting a small representative sample drawn from the original test respondents after a lapse of from two to seven days, provided additional useful information about the test's stability, an important component of test reliability. In effect, the interview results made it possible to compute an approximate coefficient of stability for the Poetry Test as a whole.

As is shown in Table LI, 79 per cent of the initial test selections within the various items were correctly recalled during the interviews. Eighty-five per cent of the "correct" responses and 33 per cent of the "incorrect" responses were recalled. For the items in which the interviewee, whether or not he selected the "correct" answer, was operating in the "correct" dimension, 92 per cent were correctly recalled, while for items for which the interviewee was not operating in the "correct" dimension, only 27 per cent were recalled.

TABLE LI

Relation of Accuracy of Recall to Correctness
of Answers and Correctness of Dimension

Per cent of Original Choices Accurately Recalled				
Total	Correct Answers	Incorrect Answers	Correct Dimension	Incorrect Dimension
79	85	33	92	27

On the whole then there appeared to be a high degree of item stability. This stability was higher for items answered correctly than for items answered incorrectly and highest of all for items in which the "correct" selection was made for the "correct" reasons. Conversely, the level of item stability was low for items answered incorrectly and lower still for items in which the selection did not involve the "correct" dimension.

In view of the limitations in the other kinds of support for test reliability, this evidence of item stability is important in indicating a reasonably adequate degree of reliability for the experimental Poetry Test.

Guessing is another source of measurement error which might be expected to reduce substantially the reliability of a test of this kind. The interview results, however, indicate a notable absence of "true", or purely random, guessing. Only 38 per cent of the original item responses made by the ten interviewees were judged not to

involve the use of the "correct" dimension. Of this 38 per cent, approximately three-quarters of the responses involved use of a dimension other than the "correct" one, leaving only one-quarter which appeared to be based on random guessing. This indicates that only one-quarter of 38 per cent, or less than 10 per cent of the total responses, were made on the basis of sheer chance.

This low incidence of random guessing is in line with the relatively high level of stability reported in previous paragraphs. If a student's original response is based on random guessing, it seems less likely that after a lapse of several days he will be able accurately to recall this response than if it is based on a reasoning process. Thus, the high degree of accuracy of recall or item stability which was encountered appears to indicate a low incidence of random guessing on the Poetry Test.

A low incidence of random guessing was also indicated by comments volunteered by respondents during the interviews. One of the two interviewees receiving the lowest test scores was so averse to guessing, and so loath to accept any of the alternatives provided that he marked in answers in an ineligible column on the IBM answer sheet as a gesture of protest.

It is difficult to determine accurately the degree of similarity between reasoning carried out at the time of the test response and the reasoning done during the subsequent interview, and there is a possibility that the

pressures of the one-to-one interview such as the desire to co-operate and the desire to appear reasonable produced some spurious rationalization. All that can be reported here is that for the vast majority of selections plausible reasons were given readily and coherently. This in itself argues against spurious rationalization, and it seems entirely possible that apparently unreasoned or intuitive initial responses were actually made unconsciously and raised to a conscious level as a result of the interview questions.

Weaknesses in Specific Items

In addition to providing information on the reliability and validity of the Poetry Test as a whole, the interview results serve to bring to light serious deficiencies in certain items. Some of the more obvious of these defects are illustrated by items A-5, A-10, and B-9.

In Item A-5, intended as a "Rhythm" item, and based on Tennyson's "Mariana", the correct alternative was "That held the pear to the gable wall". Because most of the students responding to the test were unaware that "pear" denotes "pear tree", the correct answer was rejected by most students because it appeared meaningless. Consequently the validity of this item is seriously in question.

Item A-10 was meant to test students' awareness of ironic language. In this poem, "Peace and Goodwill" are referred to as a "glib pair" of hypocritical opportunists.

What was intended to be the most tempting distractor, "Greed and Deceit", best suited the immediate context if not taken ironically. However, most respondents, apparently paying little heed to the context, noted the word "Christmas" in the first line of the poem and in a stock-response manner selected the "correct" answer, "Peace and Goodwill", for the "wrong" reasons, thus frustrating the intent of the question.

Another obvious defect is indicated by Item B-9. In this poem, the word kiss is used several times, first literally and then metaphorically, to suggest a certain kind of look in a boy's eyes. The question required the respondent to choose among the correct answer "kiss" (aptly metaphorical), "look" (too literal), "love" (metaphorical but banal), and "fire" (unsuitably metaphorical). Most respondents, seeing the word "kiss" twice previously in the poem, selected the "correct" answer purely because of repetition. Thus the attempt to measure sensitivity to metaphoric language was thwarted.

Like Item B-9 Item A-15, based on Frost's poem, "Nothing Gold Can Stay", failed. Although it was primarily intended to measure students' awareness of something relatively subtle--in this case the connotative value of a word in a certain context--interview results indicate that for most respondents the selection was made by applying a simple formula, in this case direct word repetition. "Gold", the original, was apparently selected not because it has

greater richness and precision of suggestive meaning than the distractors but simply because the word "gold" had appeared earlier in the poem. Here again failure on the investigator's part to anticipate the gap between the intent of an item and the way it would in fact be answered resulted in low validity and extremely low discriminating power (biserial correlation coefficients of 0.06 for Grade X students and 0.26 for Grade XII students).

A more subtle defect is illustrated by Item A-4, which was intended primarily to measure students' responsiveness to the sensory values of poetic language. In this item, based on Ransom's poem "Blue Girls", the original phrase, "all her perfections tarnished", is set in an immediate context which attempts to convey by means of highly concrete language ("Clear eyes fallen from blue") a sense of the horror of the physical decay of a beautiful woman. The distractors either failed to convey this sense ("ended," "forgotten"), or conveyed it in a weakened, non-sensuous, manner ("ruined"). From the interviews it appears that most of the students who selected the "correct" answer, "tarnished", made their selection mainly on the basis of sound. Because in this case the meaning of the word is inseparably linked with its sound the item is still judged to be reasonably valid. Nevertheless it is doubtful that it was successful to any significant degree in achieving its primary purpose, that of measuring response to a particular use of sensuous language.

Item B-12 serves as a final example of the usefulness of the interviews in either pointing out or explaining specific weaknesses in individual items. This item, based on Sandburg's poem "Jazz Fantasia", attempted to measure students' ability to respond to a mixture of the evocative qualities and the sound values of a word, with a heavy emphasis on the latter. Item analysis data showed the item to be unexpectedly difficult (a difficulty index of 0.08 for Grade X and 0.28 for Grade XII) and to have relatively low discriminating power (a biserial correlation coefficient of 0.32 for Grade X and 0.23, a lower figure, for Grade XII). Interview results indicate that the difficulty results mainly from one of the distractors, "bang", being either as effective as or more effective than the original, "batter", in terms of the total context and the author's intent. Of the four interviewees who made very high scores on the Poetry Test, three stoutly maintained and plausibly argued that "bang" is superior to "batter". This suggests that one difficulty in the construction of this kind of test is that on occasion a variant might be interpreted as superior to the original, even with a poem of widely recognized worth.

INTERVIEW RESULTS AND STUDENT RESPONSE

The purpose of the follow-up interviews is two-fold: to provide additional information about the experimental Poetry Test and to supplement the knowledge about students' response to poetic language gained indirectly through

statistical analyses of test results by direct questioning of respondents about their thought processes while responding to the various test items. The first of these purposes was considered in the previous section. The following section deals with information provided by the interviews concerning the way high school students respond to poetic language.

Common Characteristics in Student Response to Poetic Language

The observations made in this section result from an analysis of approximately eight hours of recorded interviews with ten students. Because most of these comments relate either directly or indirectly to "rationalization", it seems useful at this point to re-define this term and to examine the processes it denotes as related to the present study.

In its general use "rationalization" usually means the act of explaining or giving reasons for a given action or decision. For the purposes of this study "degree of rationalization" means the extent to which a given act of selection can be recalled, brought to a conscious level, and coherently explained, as determined by analysis of interview results. "Quality of rationalization" denotes (a) the degree of similarity between reasons actually underlying the original selection, as determined by interview results, and the reasons given during a subsequent interview, and (b) the extent to which this original complex of reasons approximates that which could be expected from a highly qualified reader of poetry. It is obviously more

difficult to measure and reach valid generalizations about "quality of rationalization" than about "degree of rationalization".

The following six observations concerning pupil response to poetic language are derived from the interview results:

1. The persons interviewed generally showed a much higher degree of rationalization than had been expected. In the main, students not only were able to recall what decision they had actually made but were readily able to give reasons for their selections (see Table LII). It was expected that with the high scorers the degree of rationalization would be considerably higher than with the low scorers. The actual difference, however, was only slightly in favour of the high scorers. Perhaps this is partially explained by the likelihood that a person attacking a complex problem in an overly simple fashion will find it easier to talk about his mental operations than a person whose mental operations, although perhaps more relevant to the problem at hand, are more complex.

2. The quality of rationalization, as defined above, was, like the degree of rationalization, generally higher than expected. To the extent that the term denotes "attacking the question from the point of view intended by the question-maker", it is possible to measure quality of rationalization by determining the extent to which interviewees reported using "correct" reasoning. As reported in

TABLE LII
Summary Percentage Figures for Interview Results
Based on Table XLVII

Totals for All Items and Interviewees (16 x 10)	Original Selection Correct	Original Selection Correctly Recalled	Original Selection Explained According to Correct Dimension
Possible Total	160	160	160
Actual Total	92	124.5	106.5
Per cent	57.5	77.9	66.7

the previous section (see Table XLVII), the "correct" approach dimension was used for approximately 70 per cent of all item responses about which interviewees were questioned.

To the extent that "quality of rationalization" denotes the ability to reconstruct and report with accuracy the actual thought processes underlying the original selection, it remains a construct difficult to assess with any degree of accuracy. To the extent that an "intuitive" or unconscious level of reasoning underlies the original selection it is not possible to compare an original act of selection with a subsequent recollected version. In other words, it is not possible to determine accurately to what extent an explanation for the original selection, valid as it may have been in terms of "correct dimension", was an accurate reconstruction of an original thought process. That many of the original acts of selection were of an intuitive nature appeared very clearly during the interviews. In many instances students who were at first unable to give a reason for their selection were subsequently able to do so as a result of a minimal amount of probing and a great deal of waiting on the part of the investigator. In spite of a lack of direct evidence it seems likely, particularly in view of the high item stability indicated by the interviews and the high degree of correspondence between the "correctness" of the original selecting and the "correctness" of the subsequent explaining, that most of the rationalization by the interviewees was of a high

"quality". This "high quality" can be conceived both in terms of conformity with the reasoning processes which the test items were intended to produce and in terms either of the recall of conscious thought processes or of the reconstruction at the conscious level of originally unconscious thought processes.

It was expected that the quality of rationalization would be substantially higher for the high scorers than for low scorers among the students interviewed. This expectation appears to be borne out by interview results.

3. There appears to be considerable variety in the ways different students approached an item. Generally, the low scorers seem to have approached every poem as if it were a piece of exposition, forcing "common sense" meaning into all items regardless of appropriateness to the context. Other students, apparently determined to find various kinds of patterns which they had been taught to look for in poetry, formed patterns where none existed. Still others, a smaller group, scorned "meaning", and, as it was typically expressed, "went by the way it sounds."

In contrast, the high scorers appear to have approached each poem and each item on its own terms, ready to match part to whole in a number of different ways and ready to use a wide variety of clues afforded by the context.

4. As was reported earlier, there appears in the interviews more evidence of the application of a faulty formula than of true, or random, guessing. Even when the

presence of guessing is indicated it usually involves elimination of one or two of the least likely distractors by some reasoning activity and then guessing among those remaining: a process not uncommon in the answering of multiple-choice items. -----

5. A "minimum context" principle appears to be operating with the students interviewed. High scorers and low scorers alike tended to use only the immediate context of the word or phrase in question and to use a bare minimum of clues from the larger context of the whole poem. Both high and low scorers seemed to be largely unaware of the total meaning of most of the poems. (It might be interjected here that this is not considered a serious weakness of the Poetry Test because of its focus on response to the language of poetry rather than on the meanings of whole poems.)

The difference between high and low scorers in their use of context appears to be that high scorers tended to use as much and only as much context as was needed in "solving the puzzle". In the few items such as A-12 and B-13 where knowledge of the total context was required the high scorers were highly aware of the total meaning of the poem. The low scorers, on the other hand, tended to use too little context where much was needed and too much where little was needed. This matter will be discussed in the following section.

6. In general the students interviewed showed little knowledge of the terminology of prosody. High scorers and low scorers alike were rarely able to name the metre in which a poem is written or to identify types of images. Not one had heard of half rhyme, or slant rhyme, and the terms "alliteration" and "onomatopoeia" were known to only two of the students interviewed. Most of the interviewees, including those who demonstrated a high level of confidence and skill in applying knowledge of the operation of poetic language, reported that they had studied poetic nomenclature in school but that they had found it "useless and boring" and that little of it had been remembered.

Surprisingly, about half of the students who were readily able to match the part to the whole according to a metrical pattern were unable even to count the number of stresses in a line. The interview results, therefore, suggest that a high degree of awareness of various types of patterns in poetic language can exist at a relatively unconscious level.

Comparison of High and Low Scorers

The students interviewed were selected so as to obtain a representative cross-section of very high scorers, average scorers, and below average scorers on the Poetry Test. (Very low scorers were omitted because it was felt that little useful information would be provided by students obtaining scores which are less than four, i.e.,

which are at the chance level.) It therefore appeared worthwhile to use interview results in constructing two composite profiles, one based on the responses of three highest scorers and one based on those of the three lowest scorers among the interviewees. High and low scorers are characterized in capsule form in Table LIII.

Characteristics of high scorers. The composite profile representing the "good poetry reader" is based on interviews with two Grade XII students whose score of 14 placed them at the 99th percentile for the total group and with one Grade X student whose score of 10 placed her at the 98th percentile for the Grade X group. These three interviews, together lasting slightly more than two hours, were examined in detail.

From this examination the following impressions and observations emerged:

1. The high scorer tended to need all of the time which was allotted for the test. Sample comments: "I would have liked more time to go over the questions"; "I had barely enough time." The high scorer tended to respond slowly and carefully to the poems comprising the items, taking "a lot of time" with the difficult ones and lingering over words and expressions.

2. The high scorer tended to regard the test as fairly difficult on the whole but with individual items varying in degree of difficulty. Although in both forms of

TABLE LIII
 Capsule Characterization of High and Low Scorers as
 Indicated by the Interview Results

Dimension	High Scorers	Low Scorers
1. Reading Rate per Item	low	high
2. Perception of Test's Difficulty	high	med-low
3. Rating of Test's Interest Level	high	low
4. Sensitivity to Sound and Suggestion of Words	high	low
5. Tendency to Treat Poetic Language as "Ordinary" Language	med-low	high
6. Efficient Use of Context	high	low
7. Perception that Different Skills were being Tested	high	low
8. Volume of Past Reading	high	low
9. Readiness to Make Affective Comments about Poems	high	low

the test items were placed in an order of increasing difficulty the high scorer tended to find the items becoming easier as he progressed. Sample comment: "At first it was impossible, but it got easier as I went along." In general, though, the high scorer's perception of the difficulty of individual items tended to concur with the item difficulty data reported in chapter IV.

3. The high scorer tended to find the test highly interesting. Sample comments: "It was interesting--sort of a game"; "I could have gone on for another hour"; "I disliked it at first; then I became involved, and it was really fascinating."

4. The high scorer tended to be highly flexible yet systematic in attacking the items. He looked first for obvious patterns and clues in the immediate context and if this procedure failed for progressively more subtle types of clues. He tended to re-read the poem or relevant portion of the poem and the alternatives several times, alternating between the two strategies. He realized that several types of skill were being measured and was responsive to sound and suggestion and pattern in words and expressions as well as to their "meaning". He knew with each item what was being measured (the figures in Table XLVII show that the three high-scorers averaged 15 out of 16 in this respect).

It is interesting to note in this connection what Kneller says of the creative man in The Art and Science of

Creativity.¹ According to Kneller, creativity is characterized by mental flexibility, intensity of interest, and variety of approach. It involves the ability to respond to things as they are, a tolerance for ambiguity, and a reluctance to force interpretations. This description of creativity applies very well in the present study to the mental stance of the high scorer in approaching an item. It accords with the view that the reading of poetry is a highly active, "creative" activity and offers strong evidence that the experimental test is successful in identifying good poetry readers.

5. The high scorer was able to make inferences about the total meaning of poems, but tended, above all else, to be distinguished by an unusually high sensitivity to the sound and suggestion of words and to the tone of individual passages.

6. The high scorer tended to be efficient in his use of context in "solving" an item: he used only what was necessary and relevant, and did not allow irrelevant distractors to affect his selection. When, however, a comprehensive context is required, as in Item B-13, he mastered this context.

7. The high scorer, although not necessarily a reader of poetry, tended either to read a great deal at

¹Kneller, The Art and Science of Creativity (New York: Holt, Rinehart, 1965), p. 64.

present or to have read a great deal in the past. He also tended to be a writer of poetry, although not regarding himself as proficient, and was likely to view the handling of words as "frustrating". The characteristic answer to "Have you written any poetry?" was "Some, but I'm really no good at it."

8. The high scorer tended to be curious about his own test performance and the original versions of the various poems. He tended to talk about the test with other students and even to try to discover the source of certain items.

9. The high scorer tended to make unsolicited affective comments on the various poems, such as: "I really liked this one"; "This was a cute one"; "I just couldn't see this one"; "The whole thing's sort of sick."

10. The high scorer tended to have an unfavourable view of the way in which poetry is taught. Sample comments: "I don't like the way it's being taught. Too many set poems"; "Teachers don't like teaching it."; "Have I enjoyed the poetry I've had in school? Not much."

Characteristics of low scorers. The following comments are based on three interviews, two with Grade X students whose score was 5 out of 16 and one with a Grade XII student whose score was 6 out of 16. These scores are slightly below the 30th percentile.

1. The low scorer had more than enough time to

complete the test. He reported needing little more than half the time provided. His response to the language of the poem tended to be rapid and superficial, and he showed little desire to re-read poems and linger over words and expressions.

2. The low scorer saw the test as between "fairly easy" and "medium" in difficulty. He was much less aware of differences in item difficulty than was the high scorer.

3. For the low scorer the test ranged from "kind of boring" to "fairly enjoyable." Sample comment: "I didn't like a lot of them because I couldn't understand them."

4. Whereas his high-scoring counterpart was flexible, the low scorer tended to be rigid in his approach to the various items. While the high scorer approached the poems on their own terms, receptive to the sound, suggestion, tone patterns, the low scorer tended to read them as he would a piece of non-poetic writing, looking only for referential meaning. In his approach the denotative, or "everyday" meaning tended to displace "poetic" kinds of meaning such as connotation and sound values approach.

This "displacement" concept is supported by the data presented in Table LIV which gives a summary of the interview responses by high and low scorers according to subtests. This summary clearly indicates that, while for the high scorers the "dimension score", an index of the respondent's ability to approach an item using the "correct"

TABLE LIV
Comparison Between Low and High Scorers' Identification of
"Correct" Dimension for Items in the Four Subtests

Source of Scores	Student 1 Test Dim. Score	Student 2 Test Dim. Score	Student 3 Test Dim. Score	Mean Test Dim. Score
<u>High Scorers</u>				
Subtest 1 Thought	4	3	2	3 3.7
Subtest 2 Suggestion	4	4	3	3.7 4
Subtest 3 Sound	3	3	3	3 3.3
Subtest 4 Rhythm	3	3.5	4	3.3 3.8
Total	14	15.5	14	13 14.8

TABLE LIV (Continued)

Source of Scores	Student 1		Student 2		Student 3		Mean	
	Test Score	Dim. Score	Test Score	Dim. Score	Test Score	Dim. Score	Test Score	Dim. Score
<u>Low Scorers</u>								
Subtest 1 Thought	2	4	1	3	2	4	1.7	3.3
Subtest 2 Suggestion	1	1.5	4	0.5	0	1	1.7	1
Subtest 3 Sound	1	1	0	1	1	1	0.7	1
Subtest 4 Rhythm	2	2	0	0	3	1.5	1.7	1.2
Total	6	8.5	5	4.5	5	7.5	5.3	6.8

dimension, is highly consistent among the four subtests, with the low scorers it is roughly three times as high for Subtest One ("Thought") as it is for the other three subtests.

Subtest One ("Thought") measures a type of comprehension which is to a fairly high degree common to both "everyday" and "poetic" language, while Subtests Two, Three, and Four ("Suggestion", "Sound", and "Rhythm") measure skills more peculiar to poetry. The discrepancy between "dimension scores" for Subtest One and "dimension scores" for the other subtests among low scorers may be due to the fact that the low scorer is basically unable to cope with poetic language. While he may have an adequate conception of the workings of non-poetic language, the low scorer appears to be characterized by an inadequate conception of the operation of poetic language. Here, apparently, is the chief difference between high and low scorers.

The only concession the low scorer appeared to make to the fact that he was reading poetry was an attempt to apply certain simple formulas having to do with rhyme, alliteration, word repetition and line length which he seemed to regard as essential to poetry.

The low scorer generally failed to recognize that several different skills were being measured and tended to use the same amount of context, usually in a superficial manner, for each item. The low scorer appeared unable to distinguish between what is relevant and what is not. His

method of attacking an item was to read the poem quickly and if possible select the alternative best fitting the referential meaning. When this approach failed and the "meaning" escaped him, he either selected "the one which sounded best" or simply guessed. Sample comments: "I chose whatever I felt fitted--I really used meaning"; "I had no special plan of attack; I used meaning first, then sound"; "I used the same approach for all. I used meaning." The low scorer tended to have a somewhat limited notion of the various ways in which a poem "means".

5. The low scorer tended not to be a habitual reader. He reported reading "some books", "some novels and short stories". He reported rarely reading or writing poetry: it is "too boring"; "too deep".

6. The low scorer failed to make voluntary affective comments about the poems comprising the items. When pressed, he tended to report disliking most of these poems finding them "boring".

7. The low scorer, like the high scorer, found fault with the way poetry is taught in school but saw the main cause of his dislike in the fact that poetry is intrinsically "boring" and "stupid", i.e., incomprehensible.

Possible Diagnostic Value of the Interviews

In addition to providing general information about the way high school students respond to poetic language and to pointing up differences between the way superior and

inferior poetry readers approach a poem, the follow-up interviews provide some indication that similar kinds of interviews could be valuable as a device for identifying and explaining specific areas of weakness either with individual students or with individual instructional programmes. As an example, one of the students who was interviewed failed to respond correctly to any of the "metre" items. The follow-up interview showed that he was only dimly aware of what was being measured, and had no idea of the concept of metrical language. Thus, this kind of interview promises to provide valuable information in planning teaching programmes for specific students and groups of students.

In the light of the results reported in the present chapter, such programmes would well be based on the assumption that much of the students' failure to understand and hence enjoy poetry results from ignorance of the way poetic language conveys meaning. These programmes would be directed to teaching students how to approach poetic language and might focus on such topics as "How metre contributes to meaning in poetry" and "The place of the sound of words in conveying meaning." There is nothing novel about such suggestions, but the Poetry Test and interview results strongly indicate the need for their fuller implementation in many English classrooms.

CONCLUDING REMARKS AND SUMMARY

Finally, it might again be noted that while the results reported in the present chapter appear interesting and promise to provide direction for further investigation, they are based on a very limited number of subjects, and hence must be viewed as suggestive rather than conclusive. However, certain generalizations can be made on the basis of the interview results:

1. Generally, there is a high degree of correspondence between correctness of response on the Poetry Test and correctness of "dimension" as reported in the interviews. This close correspondence, which is considerably higher for high scorers than for low scorers, provides additional evidence regarding the test's validity.

2. Generally, there is a high degree of correspondence between correctness of response on the Poetry Test and the correct recall of the original selection as reported in the interviews. This correspondence is higher for "correct" responses than "incorrect" responses, and higher for responses made using the "correct" dimension than for "correct" responses. It indicates a high level of item stability and hence provides one kind of evidence for the test's reliability.

3. The interview procedure as employed in this study proved to be a useful adjunct to the experimental test in providing direct evidence regarding weaknesses in

specific items. The reasons for these weaknesses, which could only be inferred from examination of item analysis data, were in many cases explained by the interviews. In this way interviews of this nature can be used in improving a test or in more accurately interpreting test results.

4. Generally the interview results indicated both a high degree of rationalization and a high quality of rationalization (see page 265). However, while the degree of rationalization was approximately equal for the high and low scorers, the quality of rationalization was considerably higher for the high scorers.

5. Different students tended to exhibit different strategies in approaching test items. Low scorers tended to use the same approach for each item while high scorers tended to vary their approach in accordance with the nature of the item. Some students appeared to attempt to use a logical reasoning process throughout, regardless of the kind of item, approaching poetic language as they would referential language, while others simply "went by their feelings", either initially or after the unsuccessful application of an inappropriate formula.

6. For high and low scorers alike, a "minimum-context" principle appeared to be operating. The difference is that high scorers used only as much context as was necessary to "solve" the item, often failing to understand the meaning of the whole poem, while low scorers, although using the same amount of context as high scorers overall,

tended to use too little for some items and too much for others.

7. Generally, the students interviewed displayed a minimal knowledge of poetic terminology.

8. The high scorers, in general, tended to read the poems slowly and intensively and to see the test as difficult but interesting. They were both systematic and flexible in their approach to the items and showed a high degree of sensitivity to the sound and suggestive qualities of individual words and phrases. They made efficient use of context. They reported doing a substantial amount of reading and often made unsolicited affective comments about the poems comprising the items.

9. The low scorers, in general, read the poems quickly and saw the test as fairly easy and quite uninteresting. They tended to approach each item in the same way and to read the poems as if written in strictly referential language with superimposed metre and rhyme. They failed to see that the test was attempting to measure separate skills. They reported not doing much reading, and rarely made affective comments about the poems comprising the items.

10. With some of the students interviewed, a total inability to cope with certain kinds of items became readily apparent. This indicates that interviews such as those used in the present study have considerable diagnostic value.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

SUMMARY OF THE RESEARCH

The present study attempted to investigate (a) the response of high school students to poetic language and (b) the usefulness of an investigatory procedure consisting primarily of an experimental poetry test which was developed in order to examine this response.

The Poetry Test attempted to measure student response using four inter-related yet reasonably discrete categories (Thought, Suggestion, Sound, and Rhythm). In essence, the test employed the "variant version" technique. In each item a portion of an original poem was deleted so as to require the respondent to select the original from four alternatives by using the context of the whole poem in any way he saw fit. In a "rhythm" item, for example, the metre of the original might be distorted in various ways in the three distractors so as to fail to fit the metrical pattern of the context.

Two forms of the Poetry Test, consisting of 16 items each (four in each category) were administered to 191 Grade X and XII students in North Vancouver, British Columbia, together with a standardized reading test. Follow-up interviews based on Poetry Test responses were held with a

representative sub-sample of 10 students.

Item analysis was conducted with the Poetry Test results in order to determine the difficulty level and discriminating power of each item and to provide detailed information about response patterns among the four alternatives in each item. Intercorrelations among the four subtests and between the Poetry Test and Reading Test were obtained to determine possible patterns of correspondence. Two-way analyses of variance were carried out to determine the degree of relationship between grade level, sex, and reading level, and performance on the Poetry Test. Principal components analysis was also carried out in an attempt to identify possible factors in response to poetic language.

The ten recorded interviews were analyzed in order to obtain information to supplement the results of the statistical analyses. The interviews focused on the reasons behind student choices within the various items. Interview results were examined in an attempt to identify certain common characteristics in student responses to poetic language and to characterize "good" and "poor" poetry readers.

SUMMARY OF THE MAIN FINDINGS

1. The evidence presented indicated that the procedure used in the study was a useful means of investigating student response to poetic language. The experimental

poetry test, although difficult to construct, was generally found to be sufficiently reliable and valid for the kind of research conducted in the present investigation, and proved capable of providing a wide range of information about the way high school students read poetry. The procedure developed for the interviews also proved satisfactory in eliciting information about a student's thought processes as he reads a poem and about the Poetry Test itself.

2. The Poetry Test was on the difficult side for the sample tested and was generally too difficult for the Grade X students. In terms of difficulty it was much more suitable for the Grade XII students. Also, the Poetry Test, in terms of mean item discriminating level, was more reliable for the Grade XII than for the Grade X students.

3. Low positive correlations were found between performance on each of the four subtests and between performance on items within each of the subtests. This indicated that what was being measured is a complex of partially discrete yet interconnected skills.

4. The three factors identified showed no clear relationship to the performance on the subtests. The factors appeared to relate to "analytical" (as opposed to "instinctive") abilities rather than to skills in response to traditionally "poetic" uses of language such as metre and imagery. This finding, together with the interview results, provides further evidence of the complexity of response to poetic language.

5. The low positive correlation of Poetry Test performance with Reading Test performance indicated that what the Poetry Test measures is to a large degree different from what is measured by standardized reading tests.

6. There was a much closer relationship between Grade Level and Poetry Test performance than between Reading Level and Poetry Test performance. Virtually no connection was found between sex and Poetry Test performance.

7. The interview results revealed that both the "degree" and "quality" of rationalization, as defined on page 249 were unexpectedly high. Students, when pressed, were generally able to give reasons for their various choices with a good deal of clarity and sense. While the "degree of rationalization" was approximately the same for the high and low scorers interviewed, the "quality of rationalization" was considerably higher for the high scorers.

8. The interview results revealed that in addition to a high "degree" and "quality" of rationalization, students answering the Poetry Test appeared to share certain other characteristics, such as a reluctance to guess, a propensity to use a minimal amount of context, and little knowledge of poetic terminology. However, these results also indicate that the response of high scorers differed from that of low scorers in a number of ways. As opposed to low scorers, high scorers read the poems comprising the items more slowly and intensively; they tended

to view the test as more difficult, but less interesting; they were more systematic yet more flexible in their approach to items; they made more efficient use of context, concentrating only on useful clues; they had read a great deal of imaginative literature; and they voluntarily made affective comments about the poems used for the items.

CONCLUSIONS RELATING TO THE INVESTIGATORY PROCEDURE

1. The success of a test such as the Poetry Test, designed primarily for research purposes, depends upon the degree to which some meaningful and important psychological variable has been identified and quantified. In the present study, this variable is "skill in responding to poetic language". In the light of what is known about the complexity of language in general and of poetic language in particular, and in the light of the exploratory nature of this study, it was concluded that the Poetry Test was reasonably successful in identifying and quantifying the variable under investigation. The testing procedure appeared sufficiently valid and reliable and sufficiently productive of useful information about student response to warrant further application. Perhaps a modified form in which only one aspect of poetic language is examined with a greater degree of precision than was possible in the present study could be developed.

2. The "variant version" approach used in the

Poetry Test in conjunction with the factor theory and the operations of factor analysis promises to furnish a practicable means of testing whether a test or test item which appears to measure a certain trait or component does so in terms of underlying constructs. Factor analysis, for example, can establish empirically whether hypothesized aspects of "response to poetic language" actually emerge in practice and how well they are measured by a particular testing procedure on a particular test.

3. The present investigatory procedure consisting of an objective test supplemented by follow-up interviews based on test responses promises to be useful in examining a student's thought processes as he reads a poem. The test and interview each provide information which can not be provided by the other and which is necessary in forming adequate interpretations and conclusions. The interview provides more directly, but less systematically and objectively, information which the test provides only indirectly, but more systematically and objectively. Variations of the present test-interview combination appear promising for future investigation of student response to literature.

4. The information provided by the test and interviews promises to be of diagnostic value in pinpointing weaknesses not only in individual students but also in teaching programmes. For example, a student who knows the terminology of prosody but fails to understand the operation

of metre could be easily identified.

5. The interview results indicate that the present investigatory procedure has potential value as a teaching as well as a testing device. The reactions of the students interviewed appeared to support the teaching use of "variant versions" advocated by such authorities as Murphy.¹ It seems that the initial matching process demanded by the test secured the student's involvement, encouraging him to share in "creating" a poem. His subsequent discussion, defense, and perhaps alteration of the original selection appeared to constitute an effective learning situation.

CONCLUSIONS RELATING TO STUDENT RESPONSE TO POETIC LANGUAGE

1. There is evidence in this study, especially in the low correlation between Reading Test results and Poetry Test results, that there is a substantial difference between response to poetic language and response to "non-poetic", or discursive language. In a sense the successful reading of poetry demands a specialized reading skill, and the results of the interviews indicate that the main reason for failing to respond adequately to poetic language is the tendency of the "poor reader" to approach poetic language as if it were merely an embellished form of discursive language (see the discussion of the Popean versus the Coleridgean view in

¹Geraldine Murphy, The Study of Literature in the High School (Toronto: Blaisdell, 1968), p. 102.

Chapter II).

If, as this study indicates, failure to derive meaning from poetry and hence failure to enjoy it stems mainly from ignorance about how poetic language operates and about the nature of the particular kind of meaning it attempts to communicate, the implication for the teacher of literature is clear: his chief task is to teach his students how to read poetic language.

2. More particularly, the study indicates that an intensive, text-centred approach is necessary to the full understanding and appreciation of poetry, and that this approach, if employed imaginatively and with discretion, need not induce strain or boredom in students. The "good readers", as characterized in the analysis of interview results, showed a tendency to linger over words and phrases in an deceptively passive state, willing, as Virginia Woolf puts it, to be the writer's accomplice. It should be noted that this conclusion regarding the importance of close reading in no way precludes other useful approaches to an imaginative or literary work, including those often termed "extrinsic". The present test, by the very nature of the matching technique employed both demanded close attention to the individual words and expressions of a poem and, for some respondents indicated that this close study could be interesting and perhaps even exciting.

3. The relatively low positive correlations among the four sets of subtest scores indicate that skill in

response to poetic language consists of skills which are, to a considerable degree, separate. Thus to some extent these skills can, with due awareness of their interdependence, be considered as separate for teaching purposes.

4. The generally low correlation between Reading Test performance and Poetry Test performance suggests that perhaps standardized reading tests such as the Nelson-Denny Test fail sufficiently to take into account the suggestive and sound values of words: values which appear most consistently in poetic language. The results of this study suggest that standardized reading tests, in measuring student's understanding of word meanings, need to concern themselves with a wider range of meaning than they do at present.

5. The detailed examination of item responses presented in Chapter IV reveals that the intent of many items, all of which had been examined by a committee of judges well versed in the theory and teaching of poetry, failed to coincide with response patterns. This difference between the way a skilled poetry reader responds to an item and the response of a large number of the students examined in this study underscores the need for further studies which focus on the way students actually respond to poetic language.

6. The increase of Grade XII Poetry Test means over Grade X Poetry Test means, especially for Form B, the more reliable of the two test forms, was larger than the

corresponding increase in Reading Test means. This suggests that during the high school years there is a substantial gain in students' awareness of the nuances of language which are found in their most concentrated form in poetry. To what extent this gain is due to effective poetry teaching and to what extent it is due to increased experience with language in general and to other factors is a matter for further study.

7. Generally, the interview results revealed that high scorers and low scorers alike found fault with the way poetry is currently being taught in the schools. Beyond offering the view that poetry is "boring", "dry", and "irrelevant", however, the students interviewed were generally at a loss to account for this negative attitude. Although it is easy to speculate on this matter, the present study did not attempt directly to examine student attitudes toward poetry and the teaching of poetry. This, also, is clearly a matter for further research.

IMPLICATIONS FOR THE TEACHING OF POETRY

Although the ultimate purpose of the present study is to improve the teaching of poetry by providing teachers with information about students' response to poetic language, the immediate objective is to assist in the investigation of poetic response by developing a reasonably valid and useful measuring instrument for future experimentation. By exploring a particular investigatory

procedure the study hopes to encourage the development of similar procedures.

Nevertheless, it is possible to derive from some of the study findings certain implications for the teaching of poetry and for the evaluation of "poetic understanding". Some of the more obvious of these implications, most of which have been touched on in previous sections, are presented in summary form:

1. The generally low scores on the Poetry Test, the fact that the "Thought" Subtest scores are higher for both grades than those of the other three subtests, and a number of comments by the ten students interviewed all suggest a widespread need for teaching students how to approach poetic language. This need, though difficult to demonstrate conclusively, is generally confirmed by the findings of the present study.

2. While the interview results suggest that there exists a fairly prevalent negative attitude, even among the highest scorers, toward the way in which poetry is taught, these results also suggest a higher than expected level of interest in poetry itself among young people. Again, the implication is that there is a need for more imaginative means for securing student involvement while studying poetry and for developing in students more skill in approaching poetic language.

3. The considerable differences between the performances of the two grade levels, especially with regard to

Factor B (see Table XLVII, Chapter IV) suggest that while there may be a relative leveling off of skill in the "every-day" use of language by the time students reach high school, there is a considerable development in the ability to respond to the more subtle linguistic values which are frequently found in poetic language. This finding appears to conflict with Abbott and Trabue's suggestion that poetry be deleted from the high school curriculum, and supports the view that while active interest in poetry itself may be lacking during the high school years, the skills underlying the capacity to appreciate poetry are undergoing a considerable transformation at that time. It may well be that these underlying linguistic skills are highly capable of development in most high school students, and that this development has been hindered in the recent past by negative societal attitudes toward poetry and by a lack of imaginative teaching approaches.

4. The interview results repeatedly indicated a lack of correspondence between knowledge of poetic nomenclature and the ability to respond skilfully to the various aspects of poetic language which were tested (see the sample type-scripts in Appendix D). This suggests that knowledge about poetic structure and about the various poetic devices is of little value without repeated practice in applying this knowledge. The "variant versions" technique used in this study appears to be a workable means for providing such practice.

5. The low correlation between performance on the "Thought" Subtest and performance on the other three subtests suggests that it is entirely possible for a student to have a good grasp of the discursive meaning of a poem and yet be seriously deficient in sensitivity to such aspects as sound and suggestion. Thus teaching and testing which are confined to "thought content" appear to be inadequate in terms of the total experience a poem seeks to convey.

6. The study findings strongly suggest that there are considerable individual differences in the way various students apprehend "poetic", or "imaginative" language. Some tend to delight in and linger over the sounds and suggestions of words and phrases, while others are aware only of their symbolic function. This finding, though merely serving to confirm what most teachers of literature are well aware of, underscores the need on the teacher's part to take these personal differences into account. How he deals with them is beyond the concern of this study.

7. The interview results indicate that most students, if asked, can bring what often appears to be an instinctive knowledge of the way words operate to a conscious level and can discuss the uses of words with relative clarity, sense, and interest. The oft-cited belief that high school students are either unable or unwilling to think consciously about the operations of language is not supported by the present study.

8. The interview results suggest that the close

scrutiny of individual words and phrases is compatible with a high level of enjoyment in reading poetry. Indeed, the degree of enjoyment for the students interviewed appeared to be directly related to an ability to read a poem intensively.

9. The factor analytic results, in conflicting with the fairly traditional organizational scheme used in the construction of the Poetry Test, suggest that the way poetry is normally broken down into its various components for teaching purposes should perhaps be replaced by a different type of organization, possibly based on "intuitive" versus "analytical" uses of language. This suggestion is of course highly speculative and demands a great deal of further study.

10. The "variant versions" technique used in this study appears to offer promise both in the teaching and testing of poetry. According to the interview results, this technique seemed to be effective in securing the interaction between reader and poem which, according to most authorities, is a central goal in poetry teaching. Also the study findings indicate that this technique may offer a practicable alternative to testing based either on the factual content of specific poems, on a mastery of poetic terminology, or on the general intelligence of the students. If such an alternative is considered desirable, as is assumed in this investigation, the viability of the present testing procedure should be of interest to teachers of poetry.

SUGGESTIONS FOR FURTHER RESEARCH

1. In order to obtain further information as to the reliability of the Poetry Test, the test should be administered to a larger sample of students, with each student answering both forms. Also, for the purpose of increasing reliability, each form test could be doubled in length and administered in two sittings.

2. The present study could be repeated with both university and high school students for the purpose of comparing the responses of the two groups.

3. The four categories developed for the Poetry Test could be modified or reduced in number in an attempt to gain additional insights into response to poetic language.

4. The testing procedure developed for the present study could be used to examine response to a single component of poetic language, such as metre, imagery, or tone. This more detailed kind of investigation would seek answers to such questions as the feelings suggested by the present study:

(a) In what way is response to rhythmical sound related to other types of sound occurring in poetry?

(b) In response to a metrical pattern, what are the relative effects on readers of line length, number of feet, regularity of flow, and suitability of metre to the poem's total meaning?

(c) Is there a strong connection, as suggested in this study, between response to "tone" and the maturity of the reader?

(d) What is the relationship between the effect of the sound of a word or expression and its suggestive or connotative meanings?

(e) To what extent and in what ways are readers affected by the connotative "colouring", or intensity, of a word like "shadows" or "tarnished", regardless of its appropriateness to its context?"

5. The testing procedure used in this study could be modified in a number of ways. The "parts" to be matched could consist of two or more lines rather than a single word or expression, though the resulting increase in the scope of response may be offset by a loss of precision in measurement. The poems comprising the items, in addition to being printed, could be read aloud to respondents to test possible effects on response. Certain key words and phrases in the context could be underlined to determine the effect of this kind of clue-ing. Respondents could be told which kind of skill each item is measuring to determine the effect of knowledge of the correct dimension on item difficulty level.

6. The effects of a training session on response to the various components of poetic language could be investigated. An experimental group could be shown the way metre, word sound, and the suggestive qualities of words

and phrases contribute to the total meaning of a poem, and its results compared with those of a control group.

7. The "variant versions" technique employed in the present test could be applied to fictional and other prose passages, deleting for testing purposes such peculiarly poetic elements as metre and rhyme and retaining such general elements of imaginative language as imagery, connotation, word sounds and tone. The results of such a test could then be correlated with results for a poetry test similar to the present one for the purpose of obtaining further information regarding the relationship between response to poetic language and response to imaginative language found outside poetry.

8. The Poetry Test used in the present study should be administered in conjunction with a test measuring knowledge of poetic terminology to determine what relationship exists between the two kinds of performance. Such knowledge may be of practical benefit to teachers of poetry.

9. The three factors tentatively identified in the present study could be further investigated through attempts to construct factor-pure tests for each of these three factors. This would appear to be a rather extensive process, involving sequences of tests, each a refinement of previous ones.

10. The relationship between response to "ordinary" or discursive language and response to poetic language needs

to be investigated. Knowledge of this relationship appears fundamental to the teaching of any literature employing imaginative language, particularly poetry. Before such investigation can be adequately undertaken, however, there is a need for more sensitive and comprehensive means for measuring response to discursive language than are offered by current tests of reading comprehension and vocabulary skills. The construction of such instruments would seem to be a most useful kind of research.

BIBLIOGRAPHY

A. Books

- Aiken, Conrad, (ed.) A Comprehensive Anthology of American Poetry. New York: Random House, 1944.
- Auden, W. H. Selected Poems. London: Faber and Faber, 1938.
- Ault, Norman (ed.) Elizabethan Lyrics. New York: Capricorn, 1960.
- Barfield, Owen. Poetic Diction. London: Faber and Faber, 1928.
- Bate, W. J. (ed.) Criticism: The Major Texts. New York: Harcourt, Brace, 1952.
- Bloom, Benjamin, et al. A Taxonomy of Educational Objectives: Part I. New York: Longmans, Green, 1956.
- Brooks, Cleanth, and R. P. Warren. Understanding Poetry. New York: Holt, Rinehart, 1960.
- Bruner, Jerome S. The Process of Education. Cambridge, Mass.: Harvard University Press, 1960.
- Burke, Kenneth. The Philosophy of Literary Form. Baton Rouge: Louisiana State University Press, 1941.
- Buros, O. K. The Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon, 1965.
- Buyssens, Eric. La Communication et L'Articulation Linguistique. Paris: Presses Universitaires de France, 1967.
-
- Chomsky, Noam. Language and Mind. New York: Harcourt, Brace, and World, 1968.
- Ciardi, John. How Does a Poem Mean. Boston: Houghton-Mifflin, 1959.
- College Entrance Examination Board. Freedom and Discipline in English. New York: Commission on Education, 1965.
- Conrad, Joseph. The Nigger of the Narcissis: Preface. New York: Vintage, 1965.
- Cooley, Wm. and Paul Lohnes. Multivariate Procedures for the Behavioral Sciences. New York: Wiley, 1962.

- Cronbach, Lee. Essentials of Psychological Testing. New York: Harper and Row, 1960.
- Daiches, David. Critical Approaches to Literature. Englewood Cliffs: Prentice-Hall, 1956.
- Davitz, Joel. The Communication of Emotional Meaning. New York: McGraw-Hill, 1964.
- Dewey, John. Democracy and Education. New York: MacMillan, 1938.
- Diltz, B. C. The Poetic Pilgrimage. Toronto: Clarke, Irwin, 1942.
- _____. Poetic Experience. Toronto: McLelland and Stewart, 1955.
- Dover, K. P. (ed.) Poetry. Toronto: Holt, Rinehart, 1964.
- Drew, Elizabeth. Directions in Modern Poetry. New York: Gordon Press, 1967.
- Ebel, Robert. Measuring Educational Achievement. Englewood Cliffs, N. J.: Prentice-Hall, 1965.
- Eliot, T. S. The Use of Poetry and the Use of Criticism. London: Faber and Faber, 1933.
- _____. Collected Poems. London: Faber and Faber, 1963.
- Empson, Wm. Seven Types of Ambiguity. New York: Meridian, 1955.
- Evans, W. H. and Jerry L. Walker. New Trends in the Teaching of English in Secondary Schools. Chicago: Rand, McNally, 1966.
- Eysenck, H. S. The Logical Basis of Factor Analysis. London: The University of London, 1967.
- Ferguson, G. A. Statistical Analysis. New York: McGraw-Hill, 1966.
- Frost, Robert. The Poems of Robert Frost. New York: Random House, 1946.
- Guilford, J. P. Psychological Measurement. New York: McGraw-Hill, 1954.
- _____. Psychometric Methods. New York: McGraw-Hill, 1954.

- Gulliksen, Harold. Theory of Mental Tests. New York: Wiley, 1950.
- Gunn, Thomas (ed.) Five American Poets. London: Faber and Faber, 1963.
- Gurrey, Percival. The Appreciation of Poetry. London: Oxford University Press, 1935.
- Guth, Hans. English Today and Tomorrow. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
- Hall, E. T. The Silent Language. Garden City: Doubleday, 1959.
- Harman, H. H. Modern Factor Analysis. Chicago: Chicago University Press, 1967.
- Hildum, Donald. (ed.) Language and Thought. Princeton: D. Van Nostrand, 1967.
- Hook, J. N. The Teaching of High School English. New York: Ronald Press, 1965.
- Housman, A. E. Collected Poems. New York: Holt, Rinehart, 1965.
- Hyman, S. E. The Armed Vision. New York: Vintage, 1955.
- Jackobsen, Roman and Morris Halle. Fundamentals of Language. 's-Gravenhage: Mouton, 1956.
- Jeffares, Norman (ed.) W. B. Yeats' Collected Poetry. Toronto: Macmillan, 1962.
- Kirk, Roger. Experimental Design in the Behavioral Sciences. Belmont, Cal.: Brooks and Cole, 1968.
- La Brant, Lou. The Teaching of Literature in the Secondary School. New York: Harcourt, Brace, 1931.
- Lado, Robert. Language Testing. London: Longmans, 1961.
- Lewis, C. S. An Experiment in Criticism. Cambridge University Press, 1961.
- Lindquist, E. F. (ed.) Educational Measurement. Washington: Banta, 1951.
- Loban, Walter. Literature and Social Sensitivity. Champaign, Ill.: N.C.T.E., 1960.

- Loban, W., M. Ryan, and J. Squire. Teaching Language and Literature. New York: Harcourt, Brace, 1961.
- McNemar, Quinn. Psychological Statistics. New York: Wiley, 1962.
- Muller, H. J. The Uses of English. New York: Holt, Rinehart, 1967.
- Murphy, Geraldine. The Study of Literature in High School. Toronto: Blaisdell, 1968.
- Ogden, C. K. and I. A. Richards. The Meaning of Meaning. New York: Harcourt, Brace, 1923.
- Palmer, G. H. (ed.) The English Works of George Herbert. Boston: Houghton-Mifflin, 1915.
- Pearson, N. H. (ed.) Decade. Middletown, Conn.: Wesleyan University Press, 1960.
- Perrine, Laurence. Sound and Sense. New York: Harcourt, Brace, 1956.
- Phenix, Philip H. Realms of Meaning. New York: McGraw-Hill, 1964.
- Province of Alberta. Senior High School Curriculum Guide. 1969.
- Purves, Alan C. Elements of Writing about a Literary Work. Champaign, Ill.: N.C.T.E., 1968.
- Richards, I. A. Principles of Literary Criticism. London: Routledge and Paul, 1926.
- _____. Practical Criticism. New York: Harcourt, Brace, 1929.
- _____. Coleridge on the Imagination. London: Paul, Trench, 1934.
- _____. Speculative Instruments. London: Routledge and Kegan Paul, 1955.
- Rollings, H. E. (ed.) Shakespeare's Works. London: Lippincott, 1944.
- Rommetveit, Ragnar. Words, Meanings, and Messages. New York: Academic Press, 1968.
- Rosenberg, Sheldon. Directions in Psycholinguistics. New York: Macmillan, 1965.

- Rosenblatt, Louise M. Literature as Exploration. New York: D. Appleton-Century, 1938.
- Russell, Bertrand. On Education. London: George Unwin, 1926.
- Scheffe, H. The Analysis of Variance. New York: Wiley, 1965.
- Schiller, J. P. I. A. Richards' Theory of Literature. New Haven: Yale University Press, 1969.
- Schorer, Mark (ed.) Criticism: The Foundations of Literary Judgment. New York: Harcourt, Brace, 1948.
- Scott, Wilbur. Five Approaches to Literary Criticism. New York: Macmillan, 1963.
- Sebeok, Thomas (ed.) Style in Language. New York: Wiley, 1960.
- Shapiro, Karl and Robert Beum. A Prosody Handbook. New York: Harper and Row, 1965.
- Smith, Nila B. The Development of Literary Taste. Champaign, Ill.: N.C.T.E., 1966.
- Smith, Reed. The Teaching of Literature. New York: American Book Co., 1935.
- Sponder, Stephen. The Making of a Poem. London: Hamish Hamilton, 1955.
- Squire, J. R. The Responses of Adolescents While Reading Four Short Stories. Champaign, Ill.: N.C.T.E., 1964.
- _____. Teaching English in the United Kingdom. Champaign, Ill.: N.C.T.E., 1969.
- Sweetkind, Morris. Teaching Poetry in the High School. New York: Macmillan, 1964.
- Tate, Allan (ed.) The Language of Poetry. Princeton: Princeton University Press, 1942.
- Thorndike, R. L. and E. Hagen. Measurement in Psychology and Education. New York: Wiley, 1961.
- Turbayne, C. M. The Myth of Metaphor. New Haven: Yale University Press, 1962.

- Tyler, Ralph (ed.) Educational Evaluation: New Role, New Means. Chicago: University of Chicago Press, 1969.
- Ullman, Stephen. Semantics. Oxford: Blackwell, 1962.
- Untermeyer, Louis (ed.) Modern British and American Poetry. New York: Harcourt, Brace, and World, 1962.
- Vygotsky, L. S. Thought and Language. Cambridge, Mass.: M.I.T. Press, 1962.
- Walsh, Chad. Doors into Poetry. Englewood Cliffs: Prentice-Hall, 1962.
- Warren, Robert P. Selected Essays. New York: Random House, 1958.
- Wellek, Rene and Austen Warren. The Theory of Literature. New York: Harcourt, Brace, 1956.
- Wheelwright, Philip. The Burning Fountain. Bloomington: Indiana University Press, 1954.
- Williams, Oscar (ed.) A Little Treasury of Modern Poetry. New York: Scribner's, 1952.
- _____. A Pocket Book of Modern Verse. New York: Washington Square Press, 1954.
- Wilson, James R. Responses of College Freshmen to Three Novels. Champaign, Ill.: N.C.T.E., 1966.
- Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.
- Wood, Dorothy. Test Construction. Columbia: Chas. E. Merrill, 1960.

B. Articles

- Bloom, Benjamin. "Some Theoretical Issues Relating to Educational Evaluation," in Tyler, Ralph (ed.), Educational Evaluation. Chicago: University of Chicago Press, 1969, p. 36.
- Broening, Angela. "The Development of Literary Taste in the Secondary School," in Smith, N. B. (ed.), The Development of Literary Taste. Champaign, Ill.: N.C.T.E., 1962, p. 42.

- Budd, Wm. "Research Designs of Potential Value in Investigating Problems in English," in Braddock, R., Research in the Teaching of English. Champaign, Ill.: N.C.T.E., 1967, p. 5.
- Carroll, H. A. "A Test of Prose Appreciation," English Journal, XX (Feb., 1932), p. 22.
- Carrol, John B. "The Contribution of Psychological Theory and Education Research to the Teaching of Foreign Languages," in Valdmanis, A. (ed.), Trends in Language Teaching. New York: McGraw-Hill, 1966, p. 74.
- Early, Margaret. "Stages of Growth in Literary Appreciation," English Journal, XLIX (1960), p. 164.
- Flanagan, J. C. "Evaluation and Instructional Outcomes," in Tyler, R. (ed.), Educational Evaluation. Chicago: University of Chicago Press, 1969, p. 316.
- Gunderson, Doris. "Flaws in Research Design," in Braddock, R. (ed.), Research in the Teaching of English. Champaign, Ill.: N.C.T.E., 1966, p. 27.
- Kaiser, H. T. "The Varimax Criteria for Analytic Rotations on Factor Analysis," Psychometrika, XXIII (1958), p. 187.
- Loban, Walter. "Evaluating Growth in the Study of Literature," English Journal, XXXVII (June, 1948), p. 277.
- Meckel, H. C. "Research on Teaching Composition and Literature," in Gage, N. L., Handbook of Research on Teaching. Chicago: Rand McNally, 1963, p. 966.
- Pope, Alexander. "An Essay on Criticism," in W. J. Bate, (ed.), Criticism: The Major Texts. New York: Harcourt, Brace, 1952.
- Rosenblatt, L. M. "The Acid Test for Literature Teaching," English Journal, XLV (February, 1956), p. 65.
- Ruhlen, Helen V. "Experiment in Testing Appreciation," English Journal, XIX (May, 1930), p. 138.

C. Tests

Abbott, A. and M. R. Trabue. "A Measure of Ability to Judge Poetry." New York: Teacher's College, Columbia University, 1925.

Logasa, Hannah and Martha Wright. "Tests for the Appreciation of Literature." Bloomington, Ill.: Public School Publishing Co., 1932.

Mason, Geoffrey and H. R. Kendrick. "Test of Poetry Judgment." Victoria, B. C.: The University of Victoria, 1969.

Nelson, M. J. and E. C. Denny. "The Nelson-Denny Reading Test." Boston: Houghton-Mifflin, 1960.

APPENDICES

- A. The Poetry Test
- B. Identification and Sources of Poems Used
- C. The Interview Schedule
- D. Sample Interview Typescripts
- E. The Judges
- F. Detailed Tables

APPENDIX A

The Poetry Test

To the Student

This is a test which tries to measure your awareness of certain ways a poem operates. In each question, one of the four alternatives--the "right" answer--belongs to the original poem. See if you can pick it out by using any clues you can find in the rest of the lines.

Don't worry if you don't completely understand any poem. The test does not ask you to give reasons for your choices; if certain alternatives just "sound better" to you, fine.

Try to avoid wild guessing, however, and try to examine the poems carefully before choosing, because some of the clues are not too obvious. It has been found that experienced poetry readers need about two minutes for the average item on this test.

Thank you for your help. Place all answers on the I.B.M. answer sheets. Do not write on the test paper.

POETRY TEST A

1. Though he believe it, no man is strong,
 He thinks to be called the fortunate,
 To bring home a wife, to live long.
- But he is defeated; let the son
 Sell the farm lest the mountain fall;
 His mother and her mother won.
- His fields are used up where the moles visit,
 The contours worn flat; if there show
 Passage for water he will miss it:
- Give up his breath, his woman, his team;
 No life to touch, though later there be
 Big fruit, eagles above the _____.

1. hills
2. river
3. sea
4. stream

2. When I play on my fiddle in Dooney,
 Folk dance like a wave of the sea;
 My cousin is priest in Kilvarnet,
 My brother in Mocharabuiee.
- I passed my brother and cousin:
 They read in their books of prayer;
 I read in my book of songs
 I bought at the Sligo fair.
- When we come to the end of time
 To Peter sitting in state,
 He will smile on the three old spirits,
 _____ through the gate;
- For the good are always the merry,
 Save by an evil chance,
 And the merry love the fiddle,
 And the merry love to dance.
- And when the folk there spy me,
 They will all come up to me,
 With 'Here is the fiddler of Dooney!'
 And dance like a wave of the sea.

1. And call them first
2. And welcome us all
3. But call me first
4. But won't let me

3. Then hate me when thou wilt! If ever, now!
Now, while the world is bent my deeds to cross.
Join with the spite of fortune, make me bow,

And do not drop in for an after-loss.
Ah, do not, when my heart hath scaped this sorrow,
Come in the rearward of a conquered woe;
Give not a windy night a rainy morrow,

To linger out a purposed overthrow.
If thou wilt leave me, do not leave me last,
When other petty griefs have done their spite,
But in the onset come. So shall I taste

At first the very worst of fortune's might;
And other strains of woe, which now seem woe,
Compared with _____ will not seem so.

1. loss of thee
2. present pain
3. scorn for thee
4. thy great love

4. Twirling your blue skirts, traveling the sward
Under the towers of your seminary,
Go listen to your teachers old and contrary
Without believing a word.

Tie the white fillets then about your lustrous hair
And think no more of what will come to pass
Than bluebirds that are walking on the grass
And chattering on the air.

Practice your beauty, blue girls, before it fail;
And I will cry with my loud lips and publish
Beauty which all our power shall never establish,
It is so frail.

For I could tell you a story which is true;
I know a lady with a terrible tongue,
Blear eyes fallen from blue,
All her perfections _____ - yet it is not long
Since she was lovelier than any of you.

1. ended
2. forgotten
3. ruined
4. tarnished

5. With blackest moss the flower-plots
 Were thickly crusted, one and all;
 The rusted nails fell from the knots
 That held .
 The broken sheds looked sad and strange:
 Unlifted was the clinking latch;
 Weeded and worn the ancient thatch
 Upon the lonely moated grange.¹
 She only said, "My life is dreary,
 He cometh not," she said;
 She said, "I am aweary, aweary,
 I would that I were dead!"

1. the pear to the gable wall
2. the pear to the rotting old wall
3. the pear tree to the cottage wall
4. the pear tree to the old wall

¹(grange - farmhouse)

6. Avenge, O Lord, they saints, whose bones
 Lie scattered on the Alpine mountains cold,
 Even them who kept thy truth so pure of old
 When all our fathers worshipped stocks and stones;
 Forget not, in thy book record their groans
 Who were thy sheep, and in their ancient fold
 Slain by the bloody Piedmontese, that rolled
 Mother and infant down the rocks. Their moans . . .

1. butchered
2. fallen
3. murdered
4. slaughtered

7. And I have felt
 A presence that disturbs me with the joy
 Of elevated thoughts; a sense sublime
 Of something far more deeply interfused,
 Whose dwelling is the ,
 And the round ocean and the living air
 And the blue sky and in the mind of man,
 A motion and a spirit that impels
 All thinking things, all objects of all thought,
 And rolls through all things.

1. fading light of setting suns
2. light of setting suns
3. fading light of sunsets
4. light of the sunset

8. He clasps the _____ with crooked hands;
Close to the sun in lonely lands,
Ringed with the azure world, he stands.

The wrinkled sea beneath him crawls;
He watches from his mountain walls,
And like a thunderbolt he falls.

1. bluff
2. crag
3. cliff
4. rock

9. I like to see it lap the miles
And lick the valleys up,
And stop to feed itself at tanks;
And then, prodigious, step

Around a pile of mountains,
And, supercilious, _____
In shanties by the _____ of roads;
And then a quarry pare

To fit its sides, and crawl between
Complaining all the while
In horrid, hooting stanza;
Then chase itself downhill

And neigh like Boanerges;
Then, punctual as a star,
Stop - docile and omnipotent -
At its own stable door.

1. peer
2. peek
3. stare
4. gaze

10. Punctually at Christmas the soft plush
Of sentiment snows down, embosoms all
The sharp and pointed shapes of venom, shawls
The hills and hides the shocking holes of this
Uneven world of want and wealth, cushions
With cozy wish like cotton-wool and cool
Arm's length interstices of caste and class
And into obese folds subtracts from sight
All truculent acts, bleeding the world white.

10. (Continued)

Punctually that glib pair, _____,
 Emerges royally to take the air,
 Collect the bows, assimilate the smiles,
 Of waiting men. It is a genial time;
 Angels, like stalactites, descend from heaven;
 Bishops distribute their own weight in words;
 Congratulate the poor on Christ-like lack;
 And the member for the constituency
 Feeds the five thousand, and has plenty back.

1. Greed and Deceit
2. Kindness and Love
3. Peace and Goodwill
4. Vanity and Pride

11. Fear no more the heat o' the sun
 Nor the furious winter's rages;
 Thou thy worldly task hast done,
 Home art gone and ta'en thy wages;
 Golden lads and girls all must,
 As chimney sweepers, come to dust.

Fear no more the frown o' the great,
 Thou art past the tyrant's stroke;
 Care no more to clothe and eat;
 To thee the _____ is as the oak:
 The sceptre, learning, physic, must
 All follow this and come to dust.

Fear no more the lightning flash
 Nor the all-dreaded thunder stone;
 Fear not slander, censure rash;
 Thou hast finished joy and moan;
 All lovers young, all lovers must
 Consign to thee, and come to dust.

1. reed
2. king
3. crown
4. rod

14. (Continued)

1. fix
2. set
3. catch
4. hold

15. Nature's first green is gold,
 Her hardest hue to hold.
 Her early leaf's a flower;
 But only so an hour.
 Then leaf subsides to leaf.
 So Eden sank to grief,
 So dawn goes down to day.
 Nothing _____ can stay.

1. gold
2. good
3. pure
4. rich

16. The lowest trees have tops, the ant her gall,
 The fly her spleen, the little sparks their heat;
 The slender hairs cast shadows, though but small,
 _____;
 Seas have their source, and so have shallow springs;
 And love is love, in beggars as in kings.

Where rivers smoothest run, deep are the fords;
 The dial stirs, yet none perceives it move;
 The firmest faith is found in fewest words;
 The turtles¹ cannot sing, and yet they love:
 True hearts have eyes and ears, no tongues to speak;
 They hear and see, and sign, and then they break.

1. And bees all have their stings although they
 be not great
2. And bees have stings, although not great
3. And bees' stings are most painful, though not
 great
4. And bees have stings, although they be not
 great

¹(turtles - doves)

POETRY TEST B

1. I have been one acquainted with the night.
 I have walked out in rain - and back in rain.
 I have outwalked the furthest city light.
- I have looked down the saddest city lane.
 I have passed by the watchman on his beat
 And _____, unwilling to explain.
- I have stood still, and stopped the sound of feet
 When far away an interrupted cry
 Came over houses from another street,
- But not to call me back or say goodbye;
 And further still, at an unearthly height,
 One luminary clock against the sky
- Proclaimed the time was neither wrong nor right.
 I have been one acquainted with the night.

1. dropped my eyes
2. tipped my hat
3. turned around
4. smiled at him

2. Others, I am not the first,
 Have willed more mischief than they durst:
 If in the breathless night I too
 Shiver now, tis nothing new.

More than I, if truth were told,
 Have stood and sweated hot and cold,
 And through their veins in ice and fire,
 _____ contended with desire.

Agued once like me were they,
 But I like them shall win my way
 Lastly to the bed of mould
 Where there's neither heat nor cold.

But from my grave across my brow
 Plays no wind of healing now,
 And fire and ice within me fight
 Beneath the suffocating night.

1. Fear
2. Guilt
3. Hate
4. Love

3.

In the Park
 With long spear lights
 Ride at each other
 Like armoured knights;
 Rush,
 Miss the mark,
 Pierce the dark,
 Dash by!
 Another two
 Try.

Staged
 In the Park
 From dusk
 To dawn,
 The tourney goes on:
 Rush,
 Miss the mark,
 Pierce the dark,
 Dash by!
 Another two
 Try.

1. Bikes
2. Boys
3. Cars
4. Steeds

4.

I dreamed of war-heroes, of wounded war-heroes,
 With just enough of their charms shot away
 To make them more handsome. The women moved nearer
 To touch their brave wounds and their hair streaked
 with gray.

I saw them in long ranks ascending the gang-planks;
 The girls with their doughnuts were cheerful and gay.
 They minded their manners and muttered their thanks;
 The Chaplain _____.

They shipped these rascallions, these sea-sick
 batallions
 To a patriotic and picturesque spot;
 They gave them new bibles and marksmen's medallions,
 Compasses, maps, and committed the lot.

A fine dust has settled on all that scrap metal -
 The heroes were packaged and sent home in parts
 To pluck at a poppy and sew on a petal
 And count the long night by the stroke of their
 hearts.

1. said to watch and to pray.
2. advised them to watch and to pray.
3. advised them to watch and pray.
4. said to watch and pray.

5. When I have fears that I may cease to be
 Before my pen has _____ brain,
 Before high-piled books, in charact'ry
 Hold like rich garnerers the full-ripened grain;
 When I behold, upon the night's starred face,
 Huge cloudy symbols of a high romance,
 And think that I may never live to trace
 Their shadows, with the magic hand of chance;
 And when I feel, fair creature of an hour!
 That I shall never look upon thee more,
 Never have a relish in the fairy power
 Of unreflecting love - then on the shore
 Of the wide world I stand alone, and think
 Till love and fame to nothingness do sink.

1. emptied all my
2. gleaned my teeming
3. drained my bursting
4. picked my freighted

6. O, pardon me, thou bleeding piece of earth,
 That I am meek and gentle with these butchers!
 Thou art the ruins of the noblest man
 That ever lived in the tide of the times.
 Woe to the hand that shed this costly blood!
 Over thy wounds now do I prophesy
 (Which, like dumb mouths do ope their ruby lips
 To beg the voice and utterance of my tongue),
 A curse shall light upon the limbs of men;
 Domestic fury and _____
 Shall cumber all the parts of Italy;
 Blood and destruction shall be so in use
 And dreadful objects so familiar
 That mothers shall but smile when they behold
 Their infants quartered with the hands of war,
 All pity choked with custom of fell deeds.

1. fierce civil war
2. wild civil strife
3. fierce violence
4. fierce civil strife

7. When God at first made Man,
 Having a glass of blessings standing by;
 Let us (said He) pour on him all we can:
 Let the world's riches, which dispersed lie,
 Contract into a span.

7. (Continued)

So strength first made a way;
Then beauty flowed, then wisdom, honour, pleasure:
When all was almost out, God made a stay,
Perceiving that alone, of all his treasure,
Rest in the bottom lay.

For if I should (said He)
Bestow this jewel also on my creature,
He would adore my gifts instead of me,
And rest in Nature, not the God of Nature:
So both should losers be.

Yet let him keep the rest;
But keep them with repining restlessness;
Let him be rich and weary, that at last,
If goodness lead him not, yet weariness

1. May toss him to my breast
2. May yet toss him to my breast
3. May toss him finally to my breast
4. May finally toss him to my patient breast

8. The rich man has his motor car,
 His country and his town estate.
 He smokes a fifty-cent cigar
 And jeers at Fate.

He frivols through the livelong day,
He knows not Poverty her pinch.
His lot seems light, his heart seems gay,
He has a cinch.

Yet though my lamp burns low and dim,
Though I must slave for livelihood -
Think you that I would change with him?

1. I wish I could.
2. You bet I would!
3. Oh, if I could!
4. I think I would!

9. Strephon kissed me in the spring,
 Robin in the fall,
 But Colin only looked at me
 And never kissed at all.

Strephon's kiss was lost in jest,
 Robin's lost in play,
 But the _____ in Colin's eyes
 Haunts me night and day.

1. fire
2. kiss
3. light
4. love

10. Keep me from going to sleep too soon.
 Come wake me up. Come any hour
 Of night. Come whistling up the road.
 Stomp on the porch. Bang on the door.
 Make me get out of bed and come

 Tell me the northern lights are on
 And make me look. Or tell me clouds
 Are doing something to the moon
 They never did before, and show me.
 See that I see. Talk to me till
 I'm half as wide awake as you
 And start to dress wondering why
 I ever went to bed at all.
 Tell me the walking is superb.
 Not only tell me but persuade me.
 You know I'm not too hard persuaded.

1. And open up the door and light a light
2. And let you in and light the kitchen light
3. And let you in and light a light
4. And let you into the light

11. Where had I heard this wind before
 Change like this to a deeper roar?
 What would it take my standing there for,
 Holding open a restive door,
 Looking down hill to a frothy shore?
 Summer was past and day was past,
 Sombre clouds in the west were massed.
 Out in the porch's sagging floor,
 Leaves got up in a _____ and hissed,
 Blindly struck at my knee and missed.
 Something sinister in the tone
 Told me my secret must be known:
 Word I was in the house alone
 Somehow must have gotten abroad,
 Word I had no one left but God.

1. whirl
2. coil
3. ring
4. twist

12. Drum on your drums, _____ on your banjoes,
 sob on the long cool winding saxophones,
 Go to it, O jazzmen.

Sling your knuckles on the bottoms of the happy
 tin pans, let your trombones ooze and go
 husha-husha-hush with the slippery sand-paper.

Moan like an autumn wind high in the lonesome
 treetops, moan soft like you wanted somebody
 terrible, cry like a racing car slipping away
 from a motorcycle cop, bang-bang! you jazzmen,
 bang altogether drums, traps, banjoes, horns,
 tin cans - make two people fight on the top of
 a stairway and scratch each other's eyes in a
 clinch tumbling down the stairs.

1. bang
2. batter
3. beat
4. strike

13. "But hear. If you stay, and the child be born,
 It will pass as your husband's with the rest,
 While, if we fly, the teeth of scorn
 Will be gleaming at us from east to west;
 And the child will come as a life despised;
 I feel an elopement is ill-advised!"

13. (Continued)

"O you realize not what it is my dear,
 To a woman! Daily and hourly alarms
 Lest the truth should out. How can I stay here
 And nightly take him into my arms!
 Come to the child no name or fame,
 Let us _____, and bear the shame."

1. stay, and face it,
2. go, and hide it,
3. stay, and hide it,
4. go, and face it,

14. The pale purple even
 Melts around thy flight;
 Like a star of Heaven
 In the broad daylight
 Thou art unseen, but yet I hear
 thy shrill delight.

Keen as are the arrows
 Of that silver sphere,
 Whose intense lamp narrows
 In the white dawn clear
 Until we hardly see - _____

All the earth and air
 With thy voice is loud,
 As, when night is bare,
 From one lonely cloud
 The moon rains out her beams, and
 Heaven is overflowed.

1. we feel it there.
2. we feel that it is there.
3. but yet we feel that it's there.
4. we only feel that it is there.

15. Let the boy try along this bayonet blade
 How cold steel is, and keen with the hunger of blood,
 Blue with all malice, like a madman's flash;
 And thinly drawn and famishing for flesh.

Lend him to stroke these blind, blunt bullet-heads
 Which long to nuzzle in the hearts of _____;
 Or give him cartridges of fine zinc teeth,
 Sharp with the sharpness of grief and death.

15. (Continued)

For his teeth seem for laughing round an apple.
There lurk no claws behind his fingers supple;
And God will grow no talons at his heels,
Nor antlers through the thickness of his curls.

1. men
2. lads
3. youths
4. boys

16. Smuggled sibilants,
Rattling pages,
A muted yawn,
Someone _____ ages away,
Faint footsteps overhead,
A clicked looseleaf,
Desultory drumming of bored pencils;
A faintly sweaty, chalky smell;
A weary, yet taut expectation
Of the noon bell.

1. droning
2. whispering
3. laughing
4. muttering

APPENDIX B

Poems Used for the Poetry Test Items and their Sources

Item

- A-1 W. H. Auden. Chorus from "Paid on Both Sides," in Selected Poems. London: Faber and Faber, 1938, p. 38.
- A-2 W. B. Yeats, "The Fiddler of Dooney," in Norman Jeffares, ed., W. B. Yeats Selected Poetry. Toronto: MacMillan, 1962, p. 36.
- A-3 Wm. Shakespeare's "Sonnet 90," in Shakespeares Works, H. E. Rollins, ed., London: J. B. Lippincott, 1944, p. 225.
- A-4 John Crowe Ransom, "Blue Girls," in Oscar Williams, ed., A Pocket Book of Modern Verse. New York: Washington Square Press, 1954, p. 350.
- A-5 Alfred Tennyson, from "Mariana," in James Stephens, ed., Victorian and Later English Poets. New York: American Book Company, 1934, p. 6.
- A-6 John Milton, from "On the Late Massacre in Piedmont," in Laurence Perrine, Sound and Sense. New York: Harcourt, Brace, and World, 1956, p. 193.
- A-7 Wm. Wordsworth, from "Lines Composed a Few Miles Above Tintern Abbey," in B. C. Diltz, ed., Poetic Experience. Toronto: McLelland and Stewart, 1955, p. 278.
- A-8 Alfred Tennyson, "The Eagle," in Laurence Perrine, Sound and Sense. New York: Harcourt, Brace, and World, 1956, p. 5.
- A-9 Emily Dickinson, "The Railway Train," in Oscar Williams, ed., A Pocket Book of Modern Verse. New York: Washington Square Press, 1952, p. 82.
- A-10 W. R. Rodgers, from "White Christmas," in Oscar Williams, ed., A Pocket Book of Modern Verse. New York: Washington Square Press, 1954, p. 507.
- A-11 Wm. Shakespeare, "Fear No More," in Laurence Perrine, Sound and Sense. New York: Harcourt, Brace, and World, 1956, p. 295.
- A-12 E. E. Cummings, "Portrait VIII," in Louis Untermeyer, Modern American Poetry. New York: Harcourt, Brace, and World, 1962, p. 473.

Item

- A-13 A. E. Housman, "I Did Not Lose My Heart," in The Collected Poems of A. E. Housman. New York: Holt, Rinehart, 1965, p. 198.
- A-14 T. S. Eliot, from "The Love Song of J. Alfred Prufrock," in The Collected Poems of T. S. Eliot. London: Faber and Faber, 1963, p. 13.
- A-15 Robert Frost, "Nothing Gold Can Stay," in The Poems of Robert Frost. New York: Random House, 1946, p. 235.
- A-16 Thomas Dyer, "A Modest Love," in Norman Ault, ed., Elizabethan Lyrics. New York: Capricorn, 1960, p. 275.
- B-1 Robert Frost, "Acquainted with the Night," in Oscar Williams, ed., A Little Treasury of Modern Poetry. New York: Scribner's, 1952, p. 134.
- B-2 A. E. Housman, "Others, I Am Not the First," in Oscar Williams, ed., A Pocket Book of Modern Verse. New York: Washington Square Press, 1954, p. 170.
- B-3 Mildred Weston, "Central Park Tourney," in C. P. Dover, ed., Poetry (Toronto: Holt, Rinehart and Winston, 1964), p. 7.
- B-4 Louis Simpson, from "The Heroes," in Thomas Gunn, ed., Five American Poets. London: Faber and Faber, 1963), p. 58.
- B-5 John Keats, "When I Have Fears," in B. C. Diltz, ed., Poetic Experience. Toronto: McLelland and Stewart, 1955, p. 185.
- B-6 Wm. Shakespeare, from Julius Caesar (II-i), G. L. Kittredge, ed. Boston: Ginn, 1939, p. 50.
- B-7 George Herbert, "The Pulley," in G. H. Palmer, ed., The English Works of George Herbert. Boston: Houghton-Mifflin, 1915, p. 149.
- B-8 F. P. Adams, "The Rich Man," in Laurence Perrine, Sound and Sense. New York: Harcourt, Brace, and World, 1956, p. 40.
- B-9 Sara Teasdale, "The Look," in Oscar Williams, ed., A Little Treasury of Modern Verse. New York: Scribner's, 1952, p. 597.

Item

- B-10 Robert Francis, from "Reveille," in N. H. Pearson, ed., Decade. Middleton, Conn.: Wesleyan University Press, 1960, p. 96.
- B-11 Robert Frost, "Bereft," in Oscar Williams, ed., A Little Treasury of Modern Poetry. New York: Scribner's, 1952, p. 138.
- B-12 Carl Sandburg, from "Jazz Fantasia," in Conrad Aiken, ed., A Comprehensive Anthology of American Poetry. New York: Random House, 1929, p. 283.
- B-13 Thomas Hardy, "In the Restaurant," in Louis Untermeyer, ed., Modern British Poetry. New York: Harcourt, Brace, and World, 1962, p. 30.
- B-14 F. B. Shelley, from "To a Skylark," in B. C. Diltz, ed., Poetic Experience. Toronto: McLelland and Stewart, 1955, p. 283.
- B-15 Wilfred Owen, "Arms and the Boy," in Oscar Williams, ed., A Pocket Book of Modern Verse. New York: Washington Square Press, 1954, p. 380.
- B-16 Xavier R. Warden, "Study Hall," unpublished poem, 1958.

APPENDIX C
Interview Schedule

I. Introductory

1. Time: 5 minutes

2. Purpose:

- (a) Put student at ease
- (b) Gather general information about students' perceptions of the test

3. Questions:

- (a) Enough time? Too much time?
- (b) How did you approach the questions? What was your method of attack, if any?
- (c) In answering these questions, did you generally go by meaning? feeling? sound of words and lines? What mainly was being measured?
- (d) Did you recall any of the poems? Did this affect your answering of the questions?
- (e) Do you read poetry for pleasure? Have you taken much of it in school? Do you like what you've taken? What kind do you like? Dislike?

II. Specific Items

1. Time: about 40 minutes

2. Purpose:

Extract information about students' thought processes in answering the questions.

3. Questions:

- (a) What alternative did you pick? (reliability of response)
- (b) Why was it chosen? (which dimension; change of dimension; more than one dimension)
- (c) (If not already volunteered) What parts of the poem did you use as clues? How did this matching process operate?

- (d) Which alternatives were immediately rejected? Why? What was your second choice? Why?
- (e) (If "Thought" item) Meaning of the poem.
- (f) (If "Suggestion," "Sound," "Metre" item) Can you name the metre? (or other poetic device?) What effect was the poet trying to achieve by using it, do you think?
- (g) (If not volunteered) Poem - liked? disliked?
- (h) (If not volunteered) Item easy? difficult?
- (i) Other (specify)

III. Conclusion

1. Time: 5 minutes

2. Purpose:

To obtain any additional information, mainly of a general nature, not obtained in Part II.

3. Questions:

- (a) Was the test measuring all one thing? Different things? If so, what?
- (b) What kind of item did you find easiest? Hardest?
- (c) Was the test hard? Easy?
- (d) What did you think of the kind of question used here?
- (e) Do you think this technique would be useful as a teaching device?
- (f) Other (specify)

APPENDIX D

Sample Interview Typescripts

Following are complete typescripts of two of the ten tape-recorded interviews. They were selected because of a desire to represent both grade levels and because both respondents were able to express their views clearly and coherently. In the first interview that respondent was a Grade X girl whose score of 10 placed her at the 98th percentile for the Grade X group and ranked her fourth among the students interviewed. In the second interview the subject was a Grade XII girl whose score of 14 placed her at the 98th percentile of all students writing the test and ranked her first among the students interviewed. "S" signifies the student; "I" the interviewer.

INTERVIEW ONE

I. Introduction

Item 1

- I: Did you find the test easy or hard?
 S: Some were quite easy, but there were a couple where you really had to think.
 I: Do you read poetry outside of school?
 S: Sometimes. Some of it is a little deep, and you don't understand it too well, but I like to try to understand it sometimes. I don't read a terrible lot. I read a lot of books.
 I: Do you write poetry at all?
 S: I'm not very good at it.
 I: But you do try it?
 S: Yes.
 I: Have you taken much in school?
 S: Well, I find that . . . I come from Alberta, actually, and we took an awful lot on metres and rhymes and things like that the last couple of years, and I know all the patterns, but when I came here I found that they don't know these things.
 I: Did you like the poetry you took in school?
 S: Well some of the poetry we've taken this year has been very nice.

- I: Do you think what you've taken in school has helped you? The theory, I mean.
 S: Well, yes. For something like this, it sure helps.
 I: I wonder if it would help you in reading poems.
 S: I can't really say.

II. Specific Items

Item 1

- I: Do you remember what you chose for the first one?
 S: Yes. It was number 1. I'm pretty sure.
 I: Yes. That's the one you chose before. Good. The main question is Did you have any reason for doing this?
 S: Yes. The poem.--I don't remember if I was thinking of the pattern but the poem is sort of sad--well, not really sad, but he's--like lonely, and if you were lonely and you were walking down the street you're not going to smile at a policeman or watchman or anything. You're going to do something so that you hope he doesn't notice you, and that would mean "drop your eyes."
 I: Okay. Were there any particular clues or any particular lines you looked at?
 S: I think I probably looked at the end--the last line in each one and compared, and I looked for internal rhyme, in case there was any internal rhyme or anything to make it fit.
 I: So you looked for pattern before looking for meaning and couldn't find any.
 S: I don't know, there wasn't an obvious one.

Item 2

- I: What about the second question?
 S: It must have been "guilt".
 I: That's the one you did choose. Can you tell me why?
 S: It seems to me the person that's talking here doesn't feel guilty for the things he does anymore. They're sort of company--I was pretty sure it wasn't "hate" or "love", because "love" just didn't go with the poem at all. And the "hate"--it wasn't really--suitable.
 I: So you rejected "hate" and "love" right away, and the toss-up was between the first two.
 S: I wasn't really sure.
 I: Well, the original was "fear".
 S: It was.
 I: That was your second choice, judging by what you've said.

S: Yes.
 I: So again you went mainly by meaning?
 S: Yes.

Item 3

I: How about question three?
 S: I'm pretty sure I answered "cars" here.
 I: Any reason?
 S: Well, it sure sounds like cars--that's one thing. That's their headlights at night. And "steeds"--it sounds an awful lot like horses, so I can't see horses having headlights. Or boys. And--"Bikes"--well--that's one I wasn't sure of because it seems to me bikes could ride around in the park quite easily--you know, with lights on, but--bikes don't go that fast. They don't speed around too much, so that left "cars".
 I: Then you used a logical process of elimination?
 S: Yes.
 I: "Bikes" was the second choice.
 S: Yes.
 I: So you were going strictly by meaning?
 S: Yes.

Item 4

I: Now what about question four?
 S: This one I think I picked by pattern.
 I: What do you mean by "pattern?"
 S: Sort of like "To touch their brave wounds with their hair streaked with gray." and I tried to find one that matched that. "Advised them to watch and to pray"--I think it was that one, number 2.
 I: Yes. That was the one you picked, and that was the original.
 S: Oh!
 I: Could you tell me a bit more about what you mean by "pattern?" Is it a pattern of meanings?
 S: Well, the meaning would have been just as well if they had "advised them to watch and pray," but "to watch and pray" kept the pattern up.
 I: Do you know what this is called? What aspect is being tested?
 S: Not sure.
 I: This is generally called "Metre". Can you label the particular metre word here?
 S: (Silence)
 I: Can you say, for example, whether its anapestic or dactylic?
 S: Well, it seems in some places to be anapestic and then it changes to be iambic. Maybe it's my

- imagination.
- I: Okay. But you picked the right one. Can you think of why the author used this metre?
- S: (Silence)
- I: Why did he use a tripping rhythm?
- S: Well, it sounds very good.
- I: Well, this is a leading question, but usually which rhythm is for more lighthearted stuff: di-dah, di-dah, di-dah, or di-dah-dah, di-dah-dah, di-dah-dah?
- S: The first one.
- I: Well, actually, it's the second that conveys lightness, and the author is using this metre in a sarcastic way. It's a very grave scene, talking about maimed men--guys with their arms shot off, going di-dah-dah, di-dah-dah, di-dah.
- S: Oh, yes! I see!
- I: But this is a difficult question. You see, the author is very bitter, and the bitterness is expressed ironically, and the metre is used ironically.
- S: You can see it. It sounds like a song, almost. Hardly a place for a song.
- I: Yes.

Item 5

- I: How about question five? Which answer did you pick?
- S: I think it was "gleaned my teeming".
- I: Yes. Do you know why you picked it?
- S: "Emptied all my" is definitely out because it's not really long enough, in the first place. It doesn't sound right. There should be another syllable or something. And "drained my bursting". "Bursting" doesn't sound right. And "picked my freighted"--the last word just didn't seem to fit too well. And the rest of the poem seemed to have the same kind of words like "teeming" and "gleaned".
- I: It seems you went by the feeling of the words here rather than just by the meaning.
- S: Yes. Because "gleaned" is a funny word, but it sounded good.
- I: Exactly. Now it isn't that you know the poem? It's quite well known.
- S: No.

Item 6

- I: What about question six?
- S: I think it was "fierce civil strife".
- I: Did you know that one?
- S: Well, actually, I think I've heard this somewhere before.

- I: It's from Julius Caesar.
 S: Oh. We took Julius Caesar this year. Is that the right answer?
 I: Yes. And that's the one you picked. Maybe your memory was working here.
 S: Well, at first I didn't recognize it, but whenever you're doing any type of multiple choice, if you have four answers, and three of them have the same word, the one that hasn't got the same word's going to be wrong. So I eliminated number 2. "Fierce violence" didn't seem long enough--it didn't have enough syllables--it just seemed to end. It was a toss-up between "civil war" and "civil strife".
 I: Why did you pick "strife" over "war"?
 S: "Strife" sounded familar; I think it was really memory partly.

Item 7

- I: Question seven?
 S: Well, here I looked at the other ones, at the other end ones, and I noticed that all three were quite short, so I said "May toss him to my breast." Is that the one I chose?
 I: Yes. And that's the original. And you didn't go by meaning?
 S: No. I don't think I ever read this one (laughs).
 I: You felt you didn't have to?
 S: No. Well, it just fit, like "So both should losers be". "May toss him to my breast"--they were all the same.
 I: Well, you saw the pattern. And you rejected the other three right away?
 S: Yes, because as soon as I saw how short the other lines were I decided he's not going to change in the last stanza.

Item 8

- I: Now how about question eight?
 S: I think I picked number 2 here.
 I: You did, and it was the right answer.
 S: He sort of wants to surprise you at the end, I think, because all of these say the same thing.
 I: So it isn't a question of ordinary meaning?
 S: No. It's to sort of surprise you. It's sort of a light poem. "You bet I would" just fits best.
 I: So you matched it for tone.
 S: Yes.
 I: Have you ever seen this poem before?
 S: No.

Item 9

- I: How about number nine?
 S: I said "love". I remember really debating between "love" and "kiss". Was I right?
 I: You did pick "love", and "kiss" was the original. Why did you reject the other two?
 S: Well, "fire" sounds like--uh--passion, and obviously if he didn't kiss here there was no passion, and "light" sort of sounds soft--more humorous or wise or something--obviously wasn't too suitable. Maybe I'm just a romantic, so I picked "love".
 I: "Kiss" was used, I think, at the metaphoric level, and "love" was meant to be straightforward way of expressing this.

Item 10

- I: How about question ten?
 S: I picked number 1, - no, number 3.
 I: Yes, it was 3. Can you give any reason. You seem to be uncertain about 1 and 3.
 S: I know I picked one that had "and light a light" at the end and when I saw the first one I thought that was it but then I realized it was the third one.
 I: Why did you prefer 3 over 1?
 S: I had to go strictly by pattern. In number 2, the word "kitchen" put me off, because the feeling of "kitchen" doesn't go with the rest of the poem.
 I: But why 3 over 1?
 S: It was pattern.
 I: Can you tell me how many beats to a line?
 S: I think four.
 I: Okay. How many beats in number 1?
 S: I think four.
 I: Do you know what these are called?
 S: Feet.
 I: What's the name of the four-foot pattern?
 S: Well, the five is pentameter, but I don't know the four.
 I: Fair enough.

Item 11

- I: What about number eleven?
 S: I picked 2.
 I: Yes. That was the original, as well. Why did you pick it?
 S: It sounds--well, here it says, "it got up" and "something" and "hissed", and "struck at my knee and missed". It sounds like a snake, and we think of a snake coiling, and a snake doesn't "ring" or "whirl" or "twist": it "coils".

I: So you went by the context, the surrounding words?
 S: Yes.
 I: Why do you think the author used these words for leaves--brought in the snake image?
 S: It seems to me he's out walking somewhere by himself alone, and it's sort of like being found out, and snake is something we see and it scares you, and he didn't want to be found out.

Item 12

I: What about number twelve?
 S: I think "batter".
 I: No. You picked "beat".
 S: I did?
 I: Yes. Why do you say "batter" now? Was it your second choice?
 S: I think so.
 I: What were you going by?
 S: (long pause) This one was really hard. I took out "strike" and "bang"--both too hard--didn't seem to fit. I think now, looking at it, I don't know if I like "batter" better or not. It seems to suggest how you play a banjo.
 I: You were going pretty well by the meaning of the words then?
 S: Yes.
 I: I was trying to test for sound.
 S: Was it "batter"?
 I: Yes.
 S: And I picked "beat". Isn't it funny? I'd pick "batter" now.

Item 13

I: What about 13?
 S: I think I picked number 3.
 I: Yes. Can you think why?
 S: Well they sound terribly guilty.
 I: Can you tell what's happening? Who's talking first?
 S: Sounds like a man talking to a woman who isn't married.
 I: You didn't realize that the man's talking in the first stage and the woman in the second?
 S: No.
 I: The right answer was 4, because he wants to ditch her and she loves him.
 S: I didn't realize this. I found this a hard one.

Item 14

- I: What about question 14?
 S: I'll have to read it again. I was looking very much at the end lines of the other two stanzas, and I realized that it couldn't have been number 1 because it was too short. I wasn't too sure whether it would start out with "but" or not. I don't think so.
 I: So you rejected number 3.
 S: I think I picked number 4.
 I: The original was number 2.
 S: Which one did I pick?
 I: You picked number 4, so you were quite right in your recollection; and number 4 has an extra beat. These final lines are supposed to have six feet, so here three extra ones are needed.
 S: Oh, I see what I did. I didn't count the extra beat in this one.
 I: But you were looking in the right direction. You just didn't look quite closely enough.
 S: Yes. I should have looked backwards farther in the line.

Item 15

- I: Number 15. Can you remember which one you picked?
 S: Yes. I definitely picked one of them over the other three. Oh, yes, I picked "lads" because the other two--because, like "blade" and "blood," and "apple" and "supple" sort of rhymed.
 I: So you went by some kind of rhyme. It's not an ordinary kind of rhyme? Did you know what it's called?
 S: No.
 I: No matter. It's sometimes called slant rhyme or half rhyme. Very few people notice this. So congratulations.

Item 16

- I: Now, the last one. Remember?
 S: Oh, yes. This was a cute one. I was wondering between "whispering" and "muttering". I was thinking that for alliteration someone whispering would be better than-- I don't remember now if I put "whispering" or "muttering".
 I: Those are the only two that appealed to you?
 S: Well, "droning" didn't sound right.
 I: You did pick "whispering". The author used "droning".
 S: Really? (laughs) I guess that's supposed to be a teacher.
 I: I think the reason is that if it were a long way off you wouldn't hear "whispering" or "muttering".

- S: Yes, but I think I was looking at it in a different way. I wasn't thinking of the teacher as talking. I thought it was one of the other people in the class.
- I: Oh, I see.
- S: If you're studying or something, in the back of your mind you notice someone muttering over there.

III. General Concluding Questions

- I: Did you feel that any of these were poor poems? Did you like or dislike any? You have to be really honest about this.
- S: Actually, some you could relate with more. Number 16 was cute because I could sit there writing a test and feel it. There were none I really didn't like. Number 1 I thought was sort of good.
- I: Do you think this kind of test, in not concentrating more on meaning, misses too much. Try to be frank.
- S: It all depends on what you're trying to get. I don't think so.
- I: What do you think of this as a learning device, quite aside from a testing device? Where you have to make choices and talk about them?
- S: When you're doing them, you're learning yourself. As you get farther into the test, you start to know what you're going to look for more.
- I: Do you think it has possibilities as a learning device where you make choices and talk about them?
- S: I think that's a really good idea. You know, talking about why someone chose something? I was lucky in being in Alberta, where we get much more in this poetry--scansion and stuff--like that really helped. We never took it, like reading poems, as much as this theory. I know at the time I didn't like it, so you should do it at the same time you're doing other things.
- I: Not an uninterrupted diet, then? But you think you need it?
- S: I think you need it. But it has to be mixed properly. The way we took it, they didn't apply it to poems very much.
- I: So you need a compromise?
- S: Yes.

INTERVIEW TWO

I. Introduction

- I: Did you have enough time?
- S: I finished it, but I would have liked more time to

go over it again, because of a few of them I wasn't too sure of.

I: So you worked fairly methodically?

S: Yes. I looked for pieces that I had seen before-- poems I'd read.

I: Were there any you had seen before?

S: There were a few--from the literature course.

I: Did you get any right because you'd seen them before?

S: (pause)

I: O.K. You can tell me later. Did you have any method of attack--any general way of approaching these questions?

S: Well, I read the poem over to see what the main topic was--the theme. Then looked at the choices and read it again, putting the word or phrase in to see if it sounded O.K. If it didn't fit the beat I'd eliminate that one. Then I'd try for the exact meaning. Like for some of them there, the words are just sort of switched around and it could have been either one-- it was mostly the way the other--uh--phrases were put. If the poet had a particular style, say, he revised his adjectives or something and you could go by that.

I: So you got a feeling about the style--you didn't go just by meaning?

S: No. Because some of them were so close.

I: Did you feel that one thing was being measured, or that separate things were being measured?

S: I thought several things were involved.

I: You mentioned metre for one thing.

S: Metre is several of them, because some of the phrases obviously didn't fit in.

I: Did you see anything else being measured?

S: Understanding, and metre, and style, and the language that was used.

I: How about the sound of words?

S: Oh, that was used in one of them, too.

I: How about the feeling of the words?

S: Once you read it over several times, one obviously had to fit in.

I: O. K. So you thought different things were being measured?

S: Oh, yes.

I: Did you feel the test was generally easy or hard?

S: At the beginning, I thought it was impossible, but as you went along it got easier.

I: Did you see that the easier questions were at the beginning and the harder ones at the end?

S: No.

I: This was what was intended.

S: I had a lot of trouble with the first one.

I: Maybe it's a question of conditioning.

S: I didn't do the test in order. If I recognized any, I did that one, then I did the ones before and after it.

- I: Fair enough. Did you find this boring, or was it interesting enough to keep you going?
- S: No. I thought it was really interesting. I could have kept going for another hour.
- I: So you enjoyed this kind of thing?
- S: Yes.
- I: I'm trying to see if there's a relation between skill and enjoyment.
- S: I think the people who wouldn't understand might just say "Forget it" but I thought it was really interesting and I wanted to find out what I was doing.

II. Specific Items

Item 1

- I: I think it's only fair for me to give you time seeing it's a week away. Any time you want the machine stopped, just signal. Do you remember which response you gave for number one?
- S: The first one, because he was unwilling to explain. "Tipping his hat" and "smiling" don't fit in.
- I: So you rejected those.
- S: Those two right away. This "turning around" could have done but--well, it just isn't something you'd do because he's walking.
- I: You thought that was the second best choice?
- S: Yes. I went by the meaning, rather than by the sounds of the words.
- I: Did you find it easy or hard?
- S: Quite easy.
- I: Were there any particular lines that helped you?
- S: Well, the phrase right after--"Unwilling to explain" and then I didn't see where he could turn around. I saw the whole thing as one continuous action.
- I: What do you think the poet is getting at?
- S: (pause) I'm not really sure. I don't really remember.

Item 2

- I: How about the second one?
- S: I think I picked "fear".
- I: You did, why?
- S: In that some stanza you've got two other sort of opposites--I forget what it's called--"hot and cold", "ice and fire". And I was just trying to figure what he would say. "Fear" fit in, but "guilt" also could, nad "hate" and "desire" could go together. "Hate" was out right away.
- I: You were looking for opposites?

- S: Yes. Then I narrowed it down to "fear" and "hate" and then I read it through again, and I think the sound of the word "fear" decided me.
- I: What was he talking about?
- S: I didn't pay attention to the other stanzas.
- I: It wasn't necessary for the question?
- S: No.
- I: Easy or hard?
- S: A bit harder than the first one, because they were so close.

Item 3

- I: How about question three?
- S: I think I chose "cars".
- I: Yes.
- S: What struck me right off was "long spear lights" and I can just picture the cars going by, driving through the darkness and the lights going ahead. I thought it could be "bikes" because they just have a single beam, and--I don't know. I had a bit of trouble in that one because "cars" is plural, "bikes" is plural and "lights" is plural so it didn't make any difference.
- I: So you were tempted by "bikes" a bit.
- S: But then, when they "rush by" it's cars, and it's something you can easily picture.
- I: What about the difficulty of this one?
- S: I don't know. I'd sort of put it with the first one.
- I: Pretty easy.
- S: Yes, quite easy.

Item 4

- I: Now, the fourth one.
- S: I think I chose number two.
- I: Yes.
- S: It was partly because of the number of beats in a phrase--I counted this one on my fingers about ten times. You were close to "said to watch and to pray", the first one, except that the language is--like, "said" and "advised"--I think "advised" fits better--the chaplain "advised"; he didn't say anything.
- I: Then you were going by meaning here?
- S: Meaning and metre both. I really liked this one.
- I: But you saw that it was testing your feeling for metre, to quite a large extent?
- S: Well, I didn't think that that was what was being tested, but I read a lot of poetry anyway, and if something's not quite right, then it bothers me.

I: With the sound?
 S: Yes.
 I: Which alternatives did you reject right away?
 S: Three and four: "advised them to watch and pray" and "said to watch and pray".
 I: Something's missing there?
 S: Yes.
 I: Can you name this metre?
 S: No.
 I: Can you count the number of beats to a line?
 S: Well (pause). No, I can't.
 I: Can you perceive why the guy's using this kind of metre? The effect of this kind of metre.
 S: No, but in the first line, "I dreamed of war heroes, of wounded war heroes," the emphasis usually comes on words that he wants emphasized.
 I: Do you think this kind of metre is common in English poetry?
 S: Probably not.
 I: It's not common. But you can't see its effect? Making the line move more rapidly--become lighter--and yet the subject is very heavy--that it's a kind of sarcasm?
 S: I noticed the sarcasm in the doughnuts but I didn't see it in the metre.

Item 5

I: How about question five?
 S: I think number two?
 I: Yes.
 S: I think I chose it because first I looked for the type of word used: "picked my freighted" was too hard--it just didn't seem to fit in. "drained my bursting" and "emptied all my"--the language was kind of out--they all kind of meant the same thing. So you couldn't go by meaning.
 I: So you went by the feeling of the language?
 S: Yes.
 I: Was it the sound of the words, or suggestion, or both?
 S: Both.
 I: Do you recognize the author? Can you guess?
 S: Oh--Keats?
 I: Yes; this is the way he operates. There's a clue, though--do you know what "gleaned" means?
 S: Something to do with harvesting?
 I: So in the next line, isn't there a very direct clue?
 S: Isn't there something about grain?
 S: Oh, yes.
 I: Did this influence your choice?
 S: No. I don't think so.

Item 6

I: Number six?
 S: Is this Shakespeare?
 I: Yes. Do you remember which choice?
 S: I think number two. Is that right?
 I: No.
 S: No?
 I: You picked number four, which happens to be the original.
 S: I had a real time trying to figure which of these it was. Because "wild civil strife"--you really get that uplifting, and you can really feel it, but "fierce" is--hard--and you can feel the anger too and I really chose between these two. And it's strange that I chose the opposite one this time.
 I: But you chose the right one originally.
 S: I can't remember exactly what I was thinking.
 I: You saw this one as being quite tough?
 S: Yes. I went over this one ten or twelve times.

Item 7

I: How about question seven?
 S: I picked number one.
 I: You did.
 S: I think I went by the last line in each stanza--the number of beats in it. The other ones were all too long and--for the first and second, I may have wavered a bit.
 I: What was your second choice?
 S: The second one--there was only one extra syllable. I didn't really have much trouble with that one.
 I: Did you notice the meaning of the poem?
 S: Uh--no--it didn't matter to the question.
 I: Did you like this poem?
 S: Not much. My favourites so far are the Keats and Shakespeare ones.

Item 8

I: What about question eight?
 S: I think I chose number two.
 I: You did.
 S: I went through it and I read it and thought it was kind of funny. But funny in a sad kind of way. "You bet I would"--it just seemed he was going to grab at the chance. "Oh if I could" was too longing and "I wish I could" just didn't seem like the way he would say it. "I think I would"--well he seems to have really made up his mind.

I: Can you see what I was chiefly trying to measure here?
 S: (pause)
 I: I was after tone. You talked about "funny" and "sad"--
 I think you were using "tone" terms.
 S: Yes.
 I: Easy or hard? Like it or dislike it?
 S: I liked it because I can see--he wants a change and
 he doesn't really know, he thinks he knows what he'd
 expect, if he was poor.

Item 9

I: What about nine?
 S: Number two.
 I: Right.
 S: It was the repetition of the word "kiss" and "kissed"
 kind of gave me a hint. Colin doesn't seem like he's
 going to be very--he--wild or fiery--he just isn't
 that aggressive. "Light"--that's more understanding
 and not--love or anything. And "love"--that just
 didn't seem to fit in. Too obvious. I think it was
 mainly the repetition.
 I: But you didn't go only by that. You eliminated the
 other ones for other reasons?
 S: Yes.
 I: So you really attacked it in two ways?
 S: Yes.

Item 10

I: What about ten?
 S: Number two?
 I: Yes.
 S: For a while it was the repetition of the word "light",
 but I almost chose the first one, too, and then it
 says "let you into the light"--I think it was the "i"
 that sort of got me. And the third one didn't quite
 read right.
 I: So you were going by metre then as well as meaning?
 Did they work together?
 S: They all worked together.
 I: Again, can you say how many beats there are to a line?
 S: Well, this one was weird. Uh--four?
 I: Yes. What's it called?
 S: Uh--tetrameter?
 I: Yes. So you actually knew this?
 S: But I've never--I don't really care about it.
 I: Anything else about this one?
 S: Oh, yes--the way he broke up sentences--like if you're
 reading it, like, "Come wake me up" then a break.
 Then "Come any hour"--it sounded like night and sort

of half asleep. I really liked this one.
 I: Did you get the meaning of this?
 S: Not really.

Item 11

I: What about question eleven?
 S: I chose number two.
 I: Yes. Why?
 S: The leaves "got up in a coil and hissed." "Hissed" reminded me of a snake and the other ones didn't even fit. I just read that line.
 I: Did you see why he was doing this?
 S: No, except that there were a lot of "s" sounds throughout the whole poem.
 I: Why the snake suggestion?
 S: Partly the sound.
 I: You found this quite easy?
 S: Yes.
 I: Did you like this one?
 S: Not as much as number ten, but it was pleasant to read.

Item 12

I: What about question twelve?
 S: I think I put number 1.
 I: Yes, you did.
 S: It was between "bang" and "beat", and then I read it through and the "e" sound in beat didn't quite fit in with banjo.
 I: You were going by sound, quite a bit?
 S: Yes. And then later one, you got sort of a repetition.
 I: Were you tempted by any of the other alternatives?
 S: "Batter" and "strike", no, but "beat" I was, a little. Just because it was one syllable.
 I: Curiously enough, Sandburg, the poet, used "batter". But it's possible that "bang" is still better. I think he's trying in batter to get a kind of percussive feeling. Why did you reject "batter" so readily?
 S: It just didn't sound right. Now I read it over I could see why it's used, but I would still reject it.

Item 13

I: Now, question thirteen?
 S: I think I chose number four?
 I: You are right.
 S: I remember I had a lot of trouble with this one and I had to read it over several times to get the meaning. And then I realized it was a dialogue and in the

second part there is someone answering. But with the "going" and "staying" it was a little bit confusing. And then I decided it had to be "Going" because she decided that she couldn't possibly stay here.

- I: In which stanza is "she" talking? Or both?
 S: The second one.
 I: Who's talking in the first one?
 S: The person whose child it is.
 I: Yes. What relation between these people?
 S: They're lovers, but they're not married. And the woman is married.
 I: Why is he talking this way in the first stanza?
 S: Because he doesn't want to get involved. He wants to get out of it.
 I: Yes. Why does she talk this way?
 S: Because she doesn't really love her husband and she sees the whole thing as a ticket out.
 I: You're not sure if she loves this guy?
 S: No. I don't think she does. She might pretend to herself she does. She's just trying to get away.
 I: Are you left in this poem without really knowing?
 S: Yes. I think he'll reject her in the end anyway.
 I: This is Thomas Hardy.
 S: It's quite apt.
 I: Did you find this one a bit more difficult?
 S: About average.

Item 14

- I: How about fourteen?
 S: I chose number two?
 I: Yes.
 S: I think it was the number of beats.
 I: That's the original. Did you find any difficulty here?
 S: (pause) No. This one I read a few times for the number of beats.
 I: Did you get it because you knew the poem?
 S: No. I knew the poem, but I didn't know the exact one, and I had to go through it several times, putting them all in. Like I never actually remembered the words.
 I: So you found this fairly tough?
 S: Yes. I think it was harder because I remembered the poem because I said "Oh if I get it wrong I'll shoot myself."

Item 15

- I: How about fifteen?
 S: I think I chose number two.

I: Yes. Which was that?
 S: "Lads." Because always through it there was "blade" and "blood" "flash" and "flesh".
 I: What do you call this?
 S: It wasn't a rhyme. An exact rhyme.
 I: You don't know this is called half-rhyme?
 S: No.

Item 16

I: What about sixteen?
 S: I think it was number one, "droning".
 I: Yes. Why?
 S: Because I just pictured a classroom and then the teacher just droning on and on and it's like a bee; you kind of get the--someone droning away--it just fitted in. It couldn't be "laughing" because that would mean children but "muttering"--it could have been--not really "whispering" didn't fit in.
 I: So you went by sound? Suggestion?
 S: Yes.
 I: Not just sound, eh?
 S: No. First you had to figure out who was going to be ages away.
 I: You've been in this situation?
 S: Yes.

III. General Concluding Questions

I: Any of these poems turn you off?
 S: Number 15. Kind of a gory image. "Let the boy try along this bayonet blade how cold steel is." You get a feeling of repulsion. But it was a good poem.
 I: You know Wilfred Owen?
 S: No.
 I: Anything else you liked or disliked?
 S: The last one I thought was fun, because of the words like "sibilants".
 I: How did you feel about the test as you wrote it? Be frank.
 S: When I started I said "What's the idea?" and "How am I supposed to know what the poet was going to say?"
 I: So you felt antagonistic?
 S: Yes. But when you get into it, you really try and figure it out it's so interesting--it's fascinating--you want to really get involved. I really wanted to find out how I did and what the right answer was and how I can match up with these people.

- I: Yes. I had the same feeling, making it up. Do you see this as more or less the same as most poetry tests?
- S: No. The poetry tests we usually get are either naming the number of beats in a line and all this kind of garbage or you're just given a poem and they just say "explicate it." Or relate to the style. And "What actually happened in the poem?" This was much better in that if you can use all these things like sound and the number of beats and meaning then you'll understand what's being said--like you can watch a lot of kids in the classroom and they don't really know what they're talking about but if they can memorize what somebody else said about the poem they get a good mark.
- I: So you could see I was trying to measure ability to apply the knowledge and not just the knowledge itself?
- S: After we'd written it a lot of us were talking about it and some said "I just hated it." and I thought, "Well, why?"
- I: Well, partly because it was difficult, perhaps?
- S: Yes. And some kids just hate poetry.
- I: What do you think of its possibility as a teaching device, as opposed to a testing device? Say you had a group of five students, and you gave reasons for your choices, and argued about them?
- S: This would be better, but they'd have to have the basic knowledge in the first place.
- I: They'd have to be taught how the devices were being used?
- S: Definitely. And you don't get that now.
- I: And after that this kind of thing could be valuable?
- S: Yes.
- I: Do you read poetry on your own?
- S: Yes. And I take drama lessons and read poetry out loud. And I find it helps when you do read it out loud to yourself.
- I: Do you read novels?
- S: I used to read three or four books a week, and then this year I started really liking poetry. And I love Shakespeare.
- I: Have you taken much poetry in school?
- S: Not too much.
- I: Do you generally like the way it's handled?
- S: No. It's turning too many people against it. Too many set poems. More choice. And they rely too much on the older poets but you have to study those to realize how these new forms develop. But not too much.
- I: Have you read Ciardi's book?
- S: Yes, I have.
- I: No wonder you did well. You cheated! Do you write poetry?
- S: Sometimes. I try sometimes. But I won't correct it or go over it again.

APPENDIX E

Judges Participating in the Validation of Items

Dr. J. W. Bilsland, Professor and Assistant Chairman,
Department of English, University of Alberta, Edmonton,
Alberta.

Dr. Joel L. Gajadharsingh, Assistant Professor, Faculty of
Education, University of Lethbridge, Lethbridge,
Alberta.

Mr. Donald Giffin, English Department, Strathcona Composite
High School, Edmonton, Alberta.

Mr. Richard T. Harrison, Assistant Professor, Department
of English, University of Alberta, Edmonton, Alberta.

APPENDIX F
Detailed Tables

TABLE LV

Form A Item Data Obtained by Using
Subtest Total Scores

Item Number	Biserial Correl.	Item Rel. Index
SUBTEST ONE (Thought)		
2	.65	.31
3	.75	.37
10	.60	.29
11	.59	.28
Means	.65	.32
SUBTEST TWO (Suggestion)		
4	.48	.24
6	.55	.25
12	.58	.27
15	.33	.15
Means	.49	.26
SUBTEST THREE (Sound)		
1	.72	.31
8	.79	.39
9	.71	.35
14	.76	.19
Means	.74	.31
SUBTEST FOUR (Rhythm)		
5	.69	.32
7	.68	.33
13	.76	.32
16	.76	.37
Means	.72	.33

TABLE LVI

Form B Item Data Obtained by Using
Subtest Total Scores

Item Number	Biserial Correl.	Item Rel. Index
SUBTEST ONE (Thought)		
1	.61	.29
2	.63	.32
3	.67	.32
13	.74	.36
Means	.66	.32
SUBTEST TWO (Suggestion)		
8	.84	.41
9	.60	.29
11	.72	.34
16	.50	.25
Means	.67	.32
SUBTEST THREE (Sound)		
5	.88	.40
6	.70	.35
12	.67	.26
15	.67	.32
Means	.73	.33
SUBTEST FOUR (Rhythm)		
4	.72	.35
7	.78	.38
10	.79	.39
14	.86	.29
Means	.79	.34

TABLE LVII
Item Data for Pilot Test A

Item Number	Difficulty	Biserial Correl.	Item Rel. Index
SUBTEST 1			
2	.75	.57	.25
4	.56	.55	.27
6	.81	.37	.14
12	.41	.45	.22
15	.53	.43	.22
Means	.61	.48	.22
SUBTEST 2			
3	.53	.46	.23
8	.31	.24	.11
11	.38	.37	.18
14	.22	.33	.14
19	.25	.45	.20
Means	.34	.37	.17
SUBTEST 3			
1	.88	.38	.13
10	.78	.77	.32
13	.75	.18	.08
16	.28	.30	.14
18	.16	-.02	-.01
20	.36	.19	.09
Means	.53	.30	.12
SUBTEST 4			
5	.94	.06	.02
7	.56	.65	.32
9	.66	.39	.19
17	.22	.20	.08
Means	.59	.32	.15

TABLE LVIII
Item Data for Pilot Test B

Item Number	Difficulty	Biserial Correl.	Item Rel. Index
SUBTEST 1			
1	.57	.43	.21
7	.80	.24	.10
8	.57	.24	.12
13	.57	.24	.12
16	.43	.34	.17
Means	.59	.30	.14
SUBTEST 2			
5	.43	.34	.17
10	.47	.35	.17
14	.20	.30	.12
19	.63	.61	.29
20	.30	-.18	-.08
Means	.41	.28	.13
SUBTEST 3			
2	.43	.69	.34
6	.43	.50	.25
11	.47	.38	.19
15	.33	.12	.05
18	.23	.17	.07
Means	.38	.37	.18
SUBTEST 4			
3	.70	.65	.30
4	.60	.48	.24
9	.63	.31	.15
12	.37	.33	.16
17	.40	.56	.27
Means	.54	.47	.22

TABLE LIX

Means, Standard Deviations, and Intercorrelations for Reading
Test Scores and Poetry Subtest Scores and Totals on Form A

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1	.27**						
Subtest 2	.08	.10					
Subtest 3	.11	.26**	.00				
Subtest 4	.63**	.66**	.49**	.60**			
Poetry Total	.15	.30**	.10	.48**	.42**		
Vocab.	.30**	.36**	.07	.38**	.46**	.70**	
Comp.	.24*	.35**	.09	.47**	.48**	.94**	.91**
Reading Total							
Means	1.82	1.72	1.39	1.35	6.27	35.54	38.19 73.72
S.D.	0.98	0.86	0.92	1.03	2.25	14.81	12.49 25.19

*p < 0.05

**p < 0.01

TABLE LX

Means, Standard Deviations, and Intercorrelations for Reading
Test Scores and Poetry Subtest Scores and Totals on Form B

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1							
Subtest 2	.30**						
Subtest 3	.32**	.29**					
Subtest 4	.24*	.12	.32**				
Poetry Total	.67**	.59**	.74**	.63**			
Vocab.	.42**	.15	.24*	.44**	.46**		
Comp.	.07	.15	.25*	.32**	.31**	.56**	
Reading Total	.29**	.17	.28**	.43**	.44**	.90**	.86**
Means	2.19	1.33	1.52	1.96	7.00	35.38	37.66 73.02
S.D.	0.95	0.95	1.14	1.03	2.78	12.03	10.19 19.64

*p < 0.05

**p < 0.01

TABLE LXI

Means, Standard Deviations, and Intercorrelations for Reading Test Scores
and Poetry Subtest Scores and Totals for Grade X on Form A

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1							
Subtest 2	.19						
Subtest 3	.14	-.12					
Subtest 4	.11	.17	-.24				
Poetry Total	.70**	.58**	.32*	.51**	.49**		
Vocab.	.21	.34*	.12	.41**	.48**	.77**	
Comp.	.22	.35*	.12	.36**	.52**	.94**	.95**
Reading Total	.23	.37**	.13	.41**	.52**	.94**	.95**
Means	1.55	1.57	1.33	1.02	5.47	28.44	33.49 61.93
S.D.	0.95	0.86	0.79	0.89	1.87	11.67	13.02 23.26

*p < 0.05

**p < 0.01

TABLE LXII

Means, Standard Deviations, and Intercorrelations for Reading Test Scores
and Poetry Subtest Scores and Totals for Grade XII on Form A

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1							
Subtest 2	.28*						
Subtest 3	-.01	.27					
Subtest 4	-.08	.25	.14				
Poetry Total	.49**	.71**	.62**	.56**			
Vocab.	-.14	.15	.01	.29*	.11		
Comp.	.27	.28*	-.07	.11	.22	.47**	
Reading Total	.03	.23	-.03	.25	.18	.91**	.79**
Means	2.11	1.89	1.45	1.73	7.16	43.78	43.65 87.43
S.D.	0.93	0.83	1.03	1.05	2.31	13.75	9.21 19.85

*p < 0.05

**p < 0.01

TABLE LXIII
Means, Standard Deviations, and Intercorrelations for Reading Test Scores
and Poetry Subtest Scores and Totals for Grade X on Form B

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1							
Subtest 2	.07						
Subtest 3	.05	-.13					
Subtest 4	.04	-.09	.24	.62**			
Poetry Total	.55**	.33*	.56**	.23	.36**		
Vocab.	.30*	-.11	.29*	.02	-.00	.34*	
Comp.	-.23	-.01	.18	.14	.20	.78**	.85**
Reading Total	.03	-.07	.28*				
Means	1.98	0.94	1.18	1.71	5.78	30.05	34.91
S.D.	0.96	0.87	0.98	0.99	2.05	7.87	9.39
							14.17

*p < 0.05

**p < 0.01

TABLE LXIV

Means, Standard Deviations, and Intercorrelations for Reading Test Scores
and Poetry Subtest Scores and Totals for Grade XII on Form B

	Sub- test 1	Sub- test 2	Sub- test 3	Sub- test 4	Poetry Total	Vocab. Comp.	Reading Total
Subtest 1							
Subtest 2	.42**						
Subtest 3	.48**	.47					
Subtest 4	.36**	.14	.28*				
Poetry Total	.76**	.64**	.81**	.60**	.34*		
Vocab.	.45**	.03	.08	.54**	.39**	.63**	
Comp.	.27	.12	.21	.54**	.40**	.93**	.88**
Reading Total	.41**	.07	.15	.60**			
Means	2.41	1.71	1.85	2.20	8.22	41.10	40.61
S.D.	0.90	0.86	1.18	1.01	2.87	13.06	10.19
							20.96

*p < 0.05

**p < 0.01

TABLE LXV

Two-Way Analysis of Variance on Form A:
 Factor A. Reading Level
 Factor B. Grade

Source of Variation	SS	DF	MS	F	p
SUBTEST 1					
"A" Main Effects	1.89	2	.95	.94	.39
"B" Main Effects	4.40	1	4.40	4.39	.04
"AB" Interaction	1.42	2	.71	.71	.50
Error Term	74.26	74	1.00		
Homogeneity of Variance Test		Chi square = 8.92		p = .11	
SUBTEST 2					
"A" Main Effects	3.17	2	1.58	2.17	.12
"B" Main Effects	1.87	1	1.87	2.57	.11
"AB" Interaction	1.57	2	.78	1.08	.35
Error Term	53.99	74	.73		
Homogeneity of Variance Test		Chi square = 8.35		p = .14	
SUBTEST 3					
"A" Main Effects	88.97	2	.44	.57	.57
"B" Main Effects	6.25	1	.63	.79	.38
"AB" Interaction	67.00	2	.33	.43	.65
Error Term	58.21	74	.79		
Homogeneity of Variance Test		Chi square = 4.81		p = .44	

TABLE LXV (Continued)

Source of Variation	SS	DF	MS	F	p
SUBTEST 4					
"A" Main Effects	41.96	2	2.10	2.27	.11
"B" Main Effects	1.44	1	14.46	15.67	<.001
"AB" Interaction	20.11	2	1.01	1.09	.34
Error Term	68.29	74	.92		

Homogeneity of Variance Test Chi square = 5.70 p = .34

TOTAL TEST

"A" Main Effects	23.41	2	11.71	2.71	.07
"B" Main Effects	6.30	1	62.99	14.58	<.001
"AB" Interaction	19.96	2	9.98	2.31	.11
Error Term	319.78	74	4.32		

Scheffe's Multiple Comparisons of Main Reading Level Effects

I	J	Contrast	F	p
1	2	.71	.77	.46
1	3	1.35	2.71	.07
2	3	.64	.63	.53

Homogeneity of Variance Test Chi square = 4.62 p = .46

TABLE LXVI

Two-Way Analysis of Variance on Form B:
 Factor A. Reading Level
 Factor B. Grade

Source of Variation	SS	DF	MS	F	p
SUBTEST 1					
"A" Main Effects	78.97	2	3.95	4.70	.01
"B" Main Effects	3.30	1	3.30	3.92	.05
"AB" Interaction	82.91	2	.41	.49	.61
Error Term	66.36	74	.84		

Scheffe's Multiple Comparisons of Main Reading Level Effects

I	J	Contrast	F	p
1	2	-.03	.01	.99
1	3	.63	3.41	.04
2	3	.66	3.60	.03

Homogeneity of Variance Test Chi square = 4.85 p = .43

SUBTEST 2

"A" Main Effects	.08	2	.04	.05	.95
"B" Main Effects	1.13	1	11.32	13.45	<.001
"AB" Interaction	21.88	2	1.09	1.30	.28
Error Term	66.50	79	.84		

Homogeneity of Variance Test Chi square = 1.27 p = .94

TABLE LXVI (Continued)

Source of Variation	SS	DF	MS	F	p
SUBTEST 3					
"A" Main Effects	4.21	2	2.11	1.73	.18
"B" Main Effects	6.46	1	6.46	5.31	.02
"AB" Interaction	.28	1	.14	.12	.89
Error Term	96.11	79	1.22		

Homogeneity of Variance Test Chi square = 7.13 p = .21

SUBTEST 4

"A" Main Effects	89.80	2	4.49	4.95	.01
"B" Main Effects	2.62	1	2.62	2.89	.09
"AB" Interaction	77.19	2	3.86	4.25	.02
Error Term	71.71	79	.91		

Scheffe's Multiple Comparisons of Main Reading Level Effects

I	J	Contrast	F	p
1	2	.32	.79	.46
1	3	.78	4.89	.01
2	3	.46	1.65	.20

Homogeneity of Variance Test Chi square = 3.86 p = .57

TABLE LXVI (Continued)

Source of Variation	SS	DF	MS	F	p
TOTAL TEST					
"A" Main Effects	48.59	2	24.29	4.05	.02
"B" Main Effects	9.32	1	93.18	15.54	<0.001
"AB" Interaction	19.68	2	9.84	1.64	.20
Error Term	473.55	79	5.99		
Scheffe's Multiple Comparisons of Main Reading Level Effects					
I	J	Contrast	F	p	
1	2	.45	.23	.79	
1	3	1.76	3.76	.03	
2	3	1.32	2.02	.14	
Homogeneity of Variance Test			Chi square = 10.38 p = .07		

TABLE LXVII

Two-Way Analysis of Variance on Form A:
 Factor A. Vocabulary Level
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	P
SUBTEST 1					
"A" Main Effects	.87	2	.43	.43	.65
"B" Main Effects	4.74	1	4.74	4.67	.03
"AB" Interaction	1.43	2	.72	.71	.50
Error Term	75.08	74	1.01		
Homogeneity of Variance Test				Chi square = 12.04	p = .03
SUBTEST 2					
"A" Main Effects	2.79	2	1.40	1.95	.15
"B" Main Effects	2.06	1	2.06	2.88	.09
"AB" Interaction	2.76	2	1.38	1.93	.15
Error Term	52.98	74	.72		
Homogeneity of Variance Test				Chi square = 5.99	p = .31
SUBTEST 3					
"A" Main Effects	.08	2	.04	.05	.95
"B" Main Effects	.68	1	.68	.88	.35
"AB" Interaction	2.62	2	1.31	1.70	.19
Error Term	57.10	74	.77		
Homogeneity of Variance Test				Chi square = 5.48	p = .36

TABLE LXVII (Continued)

Source of Variation	SS	DF	MS	F	p
SUBTEST 4					
"A" Main Effects	7.82	2	3.91	4.47	.01
"B" Main Effects	14.10	1	14.10	16.10	<.001
"AB" Interaction	1.79	2	.89	1.02	.37
Error Term	64.79	74	.88		

Scheffe's Multiple Comparisons of Main Vocabulary Level Effects

I	J	Contrast	F	p
1	2	.66	3.36	.04
1	3	.66	3.31	.04
2	3	.00	.00	1.00

Homogeneity of Variance Test	Chi square = 3.50	p = .62

TOTAL TEST

"A" Main Effects	16.93	2	8.47	1.86	.16
"B" Main Effects	65.21	1	65.21	14.35	<.001
"AB" Interaction	10.02	2	5.01	1.10	.34
Error Term	336.33	74	4.55		

Homogeneity of Variance Test	Chi square = 4.78	p = .44

TABLE LXVIII

Two-Way Analysis of Variance on Form B:
 Factor A. Vocabulary Level
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	p
SUBTEST 1					
"A" Main Effects	12.01	2	6.01	7.59	<.001
"B" Main Effects	3.34	1	3.34	4.22	.04
"AB" Interaction	.33	2	.17	.21	.81
Error Term	62.53	79	.79		

Scheffe's Multiple Comparisons of Main
 Vocabulary Level Effects

I	J	Contrast	F	p
1	2	.48	1.98	.14
1	3	.91	7.59	<.001
2	3	.43	1.69	.19

Homogeneity of Variance Test Chi square = 3.11 p = .68

SUBTEST 2

"A" Main Effects	.98	2	.49	.57	.57
"B" Main Effects	11.12	1	11.12	13.01	<.001
"AB" Interaction	.27	2	.14	.16	.85
Error Term	67.50	79	.85		

Homogeneity of Variance Test Chi square = 2.86 p = .72

TABLE LXVIII (Continued)

Source of Variation	SS	DF	MS	F	p
SUBTEST 3					
"A" Main Effects	3.23	2	1.62	1.32	.27
"B" Main Effects	6.07	1	6.07	4.96	.03
"AB" Interaction	.66	2	.33	.27	.76
Error Term	96.68	79	1.22		

Homogeneity of Variance Test Chi square = 2.73 p = .74

SUBTEST 4

"A" Main Effects	13.06	2	6.53	7.24	<.001
"B" Main Effects	2.52	1	2.52	2.79	.10
"AB" Interaction	4.42	2	2.21	2.45	.09
Error Term	71.24	79	.90		

Scheffe's Multiple Comparisons of Main
Vocabulary Level Effects

I	J	Contrast	F	p
1	2	.07	.03	.97
1	3	.85	5.82	<.001
2	3	.79	4.85	.01

Homogeneity of Variance Test Chi square = 8.41 p = .13

TABLE LXVIII (Continued)

Source of Variation	SS	DF	MS	F	p
TOTAL TEST					
"A" Main Effects	54.20	2	27.10	4.42	.02
"B" Main Effects	91.36	1	91.36	14.91	<.001
"AB" Interaction	2.61	2	1.30	.21	.81
Error Term	484.02	79	6.13		
Scheffe's Multiple Comparisons of Main Vocabulary Level Effects					
I	J	Contrast	F	p	
1	2	.21	.05	.95	
1	3	1.77	3.69	.03	
2	3	1.56	2.80	.07	
Homogeneity of Variance Test			Chi square = 8.41	p = .13	

TABLE LXIX

Two-Way Analysis of Variance on Form A:
 Factor A. Comprehension Level
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	P
SUBTEST 1					
"A" Main Effects	7.85	2	3.93	4.27	.02
"B" Main Effects	4.92	1	4.92	5.35	.02
"AB" Interaction	2.15	2	1.08	1.17	.32
Error Term	68.02	74	.91		

Scheffe's Multiple Comparisons of Main
 Comprehension Level Effects

I	J	Contrast	F	P
1	2	.73	3.83	.03
1	3	.57	2.40	.10
2	3	-.16	.18	.84

Homogeneity of Variance Test Chi square = 8.21 p = .14

SUBTEST 2

"A" Main Effects	5.12	2	2.56	3.52	.03
"B" Main Effects	2.02	1	2.02	2.78	.10
"AB" Interaction	.27	2	.13	.18	.83
Error Term	53.84	74	.73		

TABLE LXIX (Continued)

Source of Variation	SS	DF	MS	F	p

SUBTEST 2 (Continued)

Scheffe's Multiple Comparisons of Main
Comprehension Level Effects

I	J	Contrast	F	p
1	2	.41	1.50	.23
1	3	.61	3.39	.04
2	3	.20	.36	.70

Homogeneity of Variance Test Chi square = 3.48 p = .62

SUBTEST 3

"A" Main Effects	.19	2	.10	.13	.88
"B" Main Effects	.71	1	.71	.95	.33
"AB" Interaction	4.13	2	2.07	2.76	.07
Error Term	55.39	74	.75		

Homogeneity of Variance Test Chi square = 5.80 p = .33

SUBTEST 4

"A" Main Effects	3.35	2	1.67	1.75	.18
"B" Main Effects	14.33	1	14.33	15.01	<.001
"AB" Interaction	.30	2	.15	.16	.86
Error Term	70.63	74	.95		

Homogeneity of Variance Test Chi square = 3.97 p = .55

TABLE LXIX (Continued)

Source of Variation	SS	DF	MS	F	p
TOTAL TEST					
"A" Main Effects	35.67	2	17.83	4.02	.02
"B" Main Effects	66.50	1	66.50	15.00	<.001
"AB" Interaction	.59	2	.30	.07	.94
Error Term	328.05	74	4.43		
Scheffe's Multiple Comparisons of Main Comprehension Level Effects					
I	J	Contrast	F	P	
1	2	1.12	1.88	.16	
1	3	1.59	3.81	.03	
2	3	.47	.32	.72	
Homogeneity of Variance Test			Chi square = 5.23	p = .39	

TABLE LXX

Two-Way Analysis of Variance on Form B
 Factor A. Comprehension Level
 Factor B. Grade Level

Source of Variation	SS	DF	MS	F	p
SUBTEST 1					
"A" Main Effects	.16	2	.08	.08	.92
"B" Main Effects	2.80	1	2.80	2.99	.09
"AB" Interaction	1.17	2	.59	.63	.54
Error Term	73.05	78	.94		
Homogeneity of Variance Test Chi square = 2.67 p = .75					
SUBTEST 2					
"A" Main Effects	.75	2	.37	.44	.65
"B" Main Effects	12.16	1	12.16	14.32	<.001
"AB" Interaction	.22	2	.11	.13	.88
Error Term	66.22	78	.85		
Homogeneity of Variance Test Chi square = .90 p = .97					
SUBTEST 3					
"A" Main Effects	3.10	2	1.55	1.28	.28
"B" Main Effects	7.07	1	7.07	5.82	.02
"AB" Interaction	2.92	2	1.46	1.20	.31
Error Term	94.71	78	1.21		
Homogeneity of Variance Test Chi square = 6.63 p = .25					

TABLE LXX (Continued)

Source of Variation	SS	DF	MS	F	p
SUBTEST 4					
"A" Main Effects	3.51	2	1.75	1.80	.17
"B" Main Effects	3.22	1	3.22	3.30	.07
"AB" Interaction	7.79	2	3.89	3.99	.02
Error Term	76.14	78	.98		
Homogeneity of Variance Test Chi square = 3.02 p = .70					
TOTAL TEST					
"A" Main Effects	7.63	2	3.82	.59	.56
"B" Main Effects	98.70	1	98.70	15.13	<.001
"AB" Interaction	26.56	2	13.28	2.04	.14
Error Term	508.81	78	6.52		
Homogeneity of Variance Test Chi square = 11.44 p = .04					

TABLE LXXI

Principal Components Solution and Varimax Rotation
Transformation Matrix on Three Dimensions for
Form A Pearson Item Intercorrelations

Unrotated Factor Matrix				
Communalities		I	II	III
1	.55	.13	-.73	.06
2	.36	.37	-.25	-.39
3	.37	.34	-.30	.41
4	.48	-.20	.03	.66
5	.29	.37	-.27	-.27
6	.18	.13	.33	.24
7	.24	.36	-.26	-.22
8	.35	.44	-.12	.39
9	.22	.38	.25	.13
10	.48	-.15	-.47	.49
11	.20	.36	.28	-.04
12	.49	.64	.17	.22
13	.31	.46	.25	.20
14	.16	.19	-.34	-.08
15	.10	-.28	-.05	.12
16	.16	.34	.19	-.07
	4.95	1.93	1.55	1.47
Transformation Matrix				
	.81	.54	.22	
	.55	-.59	-.60	
	.20	-.60	.77	

TABLE LXXII

Principal Components Solution and Varimax Rotation
Transformation Matrix on Three Dimensions for
Form B Pearson Item Intercorrelations

Unrotated Factor Matrix				
Communalities		I	II	III
1	.34	.32	-.23	-.44
2	.49	.23	.62	.22
3	.18	.36	.05	-.22
4	.25	.31	-.07	-.39
5	.34	.55	.02	-.18
6	.62	.06	-.05	-.78
7	.56	.36	-.63	.21
8	.53	.61	.40	-.08
9	.50	.51	-.25	.42
10	.30	.37	-.39	.12
11	.35	.57	.05	-.15
12	.37	.27	.55	.06
13	.16	.39	-.04	.00
14	.35	.54	.01	.25
15	.15	.33	.18	.10
16	.16	.36	-.12	.13
	5.65	2.65	1.57	1.44
Transformation Matrix				
	.69	.52	.50	
	-.53	.84	-.13	
	.49	.18	-.85	