

DOCUMENT RESUME

ED 230 576

TM 830 272

AUTHOR Gray, James; And Others  
 TITLE Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report.  
 INSTITUTION California Univ., Berkeley. Graduate School of Education.  
 SPONS AGENCY National Inst. of Education (ED), Washington, DC.  
 PUB DATE Nov 82  
 GRANT NIE-G-80-0034  
 NOTE 559p.; Produced by the Bay Area Writing Projects.  
 PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF02/PC23 Plus Postage.  
 DESCRIPTORS Data Collection; Evaluators; \*Holistic Evaluation; Longitudinal Studies; Psychometrics; \*Scoring; Secondary Education; \*Test Construction; \*Validity; \*Writing Evaluation  
 IDENTIFIERS \*Bay Area Writing Project; Writing Assessment; \*Writing Test Prompts

ABSTRACT

Five studies of holistic writing assessment procedures examined interactive relationships of the participants, processes, and products of writing assessment episodes. The first study examined practices in designing writing test prompts. The second study investigated the effects of variation in the specification of audience in a writing test prompt upon holistic score. A second part of this study examined rater response to the anchor papers used in training three groups of raters for scoring. The third study investigated writing task and response variables and validated an analysis system for studying the relationships between assessment prompts, written products, and holistic scores. The second part of the third study summarized data collected on the performance elicited on eight prompts in a county-wide writing assessment. The fourth study investigated differences in expectations among participants in a writing test episode, and included an analysis of characteristics of a prompt which influence variations in interpretation. The fifth study investigated holistic assessment as a longitudinal measure of student growth, focussing on aspects of performance variation over time. A discourse typology was developed and applied in an analysis of performance variation on annual assessment topics over a period of five years. (Author/PN)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED230576

**PROPERTIES OF WRITING TASKS: A STUDY OF  
ALTERNATIVE PROCEDURES FOR HOLISTIC  
WRITING ASSESSMENT**

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

✓ This document has been reproduced as  
received from the person or organization  
originating it.  
Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official NIE  
position or policy.

**Final Report  
NIE-G-80-0034**

**November 1982**

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

Leo P. Rath

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

**Principal Investigator  
James Gray**

**Project Director  
Leo Ruth**

**Research Coordinators  
Sandra Murphy and Catharine Keech**

**Research Associates  
Karen Carroll, Charles Kinzer, Don Lee,  
Ann Robyns, Elissa Warantz**

**Bay Area Writing Project  
Graduate School of Education  
University of California, Berkeley**

The activity which is the subject of this report was supported in whole or in part by the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TM 800 222

## ACKNOWLEDGMENTS

The Writing Assessment Project, as this research came to be known, was conducted under the sponsorship of the Bay Area Writing Project (BAWP) in the Language & Literacy Division of the School of Education, University of California, Berkeley, under a grant from the National Institute of Education. In particular, James Gray, Director of BAWP, and Miles Myers, its Associate Director, provided needed professional and personal support, as did Marcia Farr of NIE.

No educational research can be undertaken without the cooperation of people in schools who supply the necessary data. We owe an immense debt of gratitude to the many BAWP Teacher-Consultants from local schools who served as readers in the several scorings of writing samples that we undertook. We were fortunate also to have the cooperation of several S.F. Bay Area schools--Drake High School, Los Gatos High School, Northgate High School, and the Santa Clara County Consortium of Schools.

The success of research depends also upon the quality of the perceptions and the commitment to scholarship of those who conduct it. The Writing Assessment Project was unusually fortunate in attracting an assemblage of gifted young graduate student researchers, of whom several already have taken posts at leading colleges and universities. As coordinators of our research studies, Sandra Murphy and Catherine Keech each made unique contributions essential to the success of the project. As research associates, Karen Carrol, Charles Kinzer, Don Leu, and Elissa Warantz each prepared individually distinctive analyses of data. This group, together with faculty member Paul Ammon and visiting scholar Judith Langer, constituted a collegial circle of friendly critics, meeting in weekly discussions during the first year of the Project. These discussions helped enlarge the perspective of the Writing Assessment Project. Sarah Freedman, as consultant, also provided valuable advice.

Three others deserve special appreciation: Mary Ellen McNelly and Ann Robyns co-authored portions of two of the chapters; and Deborah Ruth edited major portions of the manuscript, thereby increasing the unity, coherence, and intelligibility of the whole.

On behalf of all the members of the Writing Assessment Project Team, I wish to express our gratitude to the various assistants who undertook such chores as transcribing, typing and re-typing protocols, counting features of texts, and other necessary tasks. Among these, special thanks must go to Michael J. Brady who was there to the very end. Marta Guzman, our secretary, enabled us to keep a complex project running smoothly, and other members of the BAWP staff also added their support when it was most needed.

The authors of the reports are identified in each chapter; however, I bear overall responsibility for the final shape of the Project and this report.

Leo Ruth  
Project Director  
November, 1982

## ABSTRACT

During 1980-82, five studies of holistic writing assessment procedures were conducted in several school districts in northern California. These studies sponsored by the Bay Area Writing Project under a grant from the National Institute of Education examine interactive relationships of the participants, processes, and products of writing assessment episodes. The first study examines practices in designing writing test prompts and presents a report of these practices that was reviewed by a panel of teachers, researchers, and professional test makers. This study also reviews the literature from major fields related to the testing of writing. The second study investigates the effects of variation in the specification of audience in a writing test prompt upon holistic score. A second part of this study examines rater response to the anchor papers used in training three groups of raters for scoring. The third study investigates writing task and response variables which potentially result in discrepancies in holistic scores, and validates an analysis system for studying the relationships between assessment prompts, written products, and holistic scores. The second part of the third study summarizes data collected on the performance elicited on eight prompts in a county-wide writing assessment. The fourth study investigates differences in expectations among participants in a writing test episode, and includes an analysis of characteristics of a prompt which influence variations in interpretation. The fifth study investigates holistic assessment as a longitudinal measure of student growth, focussing on aspects of performance variation over time. A discourse typology is developed and applied in an analysis of performance variation on annual assessment topics over a period of five years.

Executive Summary

PROPERTIES OF WRITING TASKS:  
A STUDY OF ALTERNATIVE PROCEDURES FOR  
HOLISTIC WRITING ASSESSMENT

Leo Ruth, Project Director

Background and Rationale .....	1
The Research Problem .....	5
Description of the Research .....	6
Theoretical Grounding of the Research .....	6
The Model of the Writing Assessment Episode .....	7
The Principal Studies .....	9
Highlights of the Studies .....	10
Study 1: Designing Prompts for Holistic Writing Assessments: Knowledge from Theory, Research, and Practice .....	10
Study 2: Effects of Variation in a Writing Test Prompt Upon Holistic Score and Other Factors .....	15
Study 3: Beyond the Field Test of Writing Prompts: A Preliminary Investigation of the Nature of Response Variation to Selected Field- Tested Prompts .....	19
Study 4: A Study of the Construction of Meanings of a Written Prompt by Its Authors, the Student Writers, and the Raters .....	21
Study 5: Holistic Assessment As Longitudinal Measure of Student Growth: Interpreting Student Task Constructions .....	24
References .....	29

# DESCRIPTIVE TABLES OF CONTENTS

## Chapter 1

### DESIGNING PROMPTS FOR HOLISTIC WRITING ASSESSMENTS: KNOWLEDGE FROM THEORY, RESEARCH, AND PRACTICE

#### Part I

##### SOURCES OF KNOWLEDGE FOR DESIGNING WRITING TEST PROMPTS

Overview of the Study .....	32
The Research Problem .....	32
Limitations of the Study .....	34
Class Writing Assignments and School Writing Tests Distinguished .....	34
Sources of Knowledge from Theory and Research .....	36
Psychometrics .....	37
Professional Wisdom .....	42
Fred N. Scott (1903) .....	43
Commission on English (1965) .....	44
Josephine Miles (1962, 1979) .....	45
Commission on Writing (1966) .....	46
William F. Irscher (1979) .....	47
Edmund Farrell ( ) .....	48
Arthur J. Carr (1965) .....	49
Professional Consensus on Principles .....	49
Marjorie Kirrie (1979) .....	50
NAEP's Primary Trait System (1976) .....	51
Discourse Theory .....	52
Alexander Bain (1890) .....	53
Leo Rockas (1964) .....	54
James Kinneavy (1971) .....	55
James Moffett (1968) .....	56
James Britton (1975) .....	58
Richard Lloyd-Jones and Carl Klaus (1974) .....	59
Skills Model Limitations .....	61
Research .....	61
Basic Research Questions .....	61
Literature Search .....	63
Inadequate Research Report Practices .....	63
Evidence from Research .....	66

Selection of the Subject and Problems of Prior Knowledge .....	66
The Wording of Prompts and the Problem of Comprehension .....	72
Selection of Mode .....	76
Specification of Rhetorical Context:	
Role, Audience, Purpose .....	78
Contextual Influences on Assessment .....	85
What Counts As Research: Procedures of the Study .....	91
Knowledge Without Numbers .....	91
A Collaborative Model for Validating the "Theory of Idealized Practice" .....	94
Catharine Keech As Informant .....	94
The Review Panels .....	95
A Recursive, Iterative Model: Theory, Practice, and Research Interacting .....	96
Results of the Study .....	98
Responses from the Panel of Writing Instructors .....	98
Value of the Study .....	99
Classification of Writing Prompts by Discourse Mode ..	99
The Content and Wording of Prompts .....	101
Purposes of Assessment and Rhetorical Purpose .....	104
The Function of Instructions .....	104
Time Constraints .....	105
Use of Pilot Tests .....	106
Responses from the Panel of Researchers and Evaluators .....	106
Value of the Study .....	107
Issue of Assessment Purpose .....	107
Need for Definitions .....	108
The Content and Wording of Prompts .....	109
Purpose and Audience .....	110
Functions of Instructions .....	110
Time Constraints .....	111
Use of Pilot Tests .....	111
Effects of Recommendations on the Final Version of the Manuscript on Practice .....	114
More Results of the Study .....	115
Philosophy of the Assignment for Writing Assessment .....	115
Guidelines for Developing Prompts for Writing Assessments .	116
Tentative Schema of Writing Tasks .....	118
Schema for Task Analysis of Prompts .....	119
References .....	120
Appendices .....	129

# Chapter 1

## Part II

### PRACTICES IN DESIGNING TEST PROMPTS: ANALYSIS AND RECOMMENDATIONS

Introduction .....	132
The Nature of the Writing Prompt .....	135
A Note on Terms .....	135
The Structure of the Prompt: Topic and Instructions .....	135
The Writing Topic: Content and Forms .....	139
Section Overview .....	139
Topics that Draw Upon Personal Experience .....	140
Cuing Personal Topics for Content .....	144
Topics that Elicit Made-up Stories .....	149
Hypothetical Situations as Subjects for Writing .....	150
The Problem of Prior Knowledge Requirements in Exposition and Argument Topics .....	155
Controversial Issues as Topics for Persuasive Writing .....	157
Topics with Hidden Biases .....	162
The Writing Instructions: Purposes and Problems .....	166
Section Overview .....	166
Hidden Agendas in Instructions .....	166
The "Describe" Trap .....	166
The Special Conventions Trap .....	170
The Pick-a-Form Trap .....	174
Helpful and Unhelpful Instructions .....	175
Suggestions that Become Constraints .....	175
The Effects of Overprompting .....	176
Specifying Number of Words, Sentences, Paragraphs .....	179
The Value of Focusing Cues .....	182
The Effects of Specifying Audience .....	183
Conclusion .....	189
References .....	190
Appendix A: .....	193
Suggestions for Writing Essay Questions (Gertrude Conlan) .....	193
Some Considerations in Judging the Worth of an Essay Topic (Gertrude Conlan) .....	196
Appendix B: .....	198
Prompt Index with Sources .....	198

Chapter 2

EFFECTS OF VARIATION IN  
A WRITING TEST PROMPT UPON  
HOLISTIC SCORE AND OTHER FACTORS

Part I

EFFECTS OF TWO VERSIONS OF A WRITING PROMPT  
UPON HOLISTIC SCORE AND WRITING PROCESSES

Overview of the Study .....	215
Procedures of the Study .....	217
Subjects .....	217
Determination of Verbal Skill and Attitude Ratings .....	217
Timed Writing Test Versions .....	218
Administration of the Test and the Holistic Rating .....	220
Administration of the Questionnaire .....	221
Choice of Statistical Tests .....	222
General Results .....	224
Verbal Skill, Attitude, and Holistic Score .....	224
Sex and Holistic Score .....	224
Proficiency Level and Holistic Score .....	224
Effect of Prompt on Holistic Score and Choice of Mode .....	225
Questionnaire Data and Commentary .....	228
Memory for the Assigned Prompt .....	228
Task Preference .....	228
Recollections About the Planning Process .....	229
Attitudes Toward the Test Subject .....	230
Writers' Self-Profiles and Other Writing Experiences .....	230
Relations Between Selected Variables and Questionnaire Items .....	233
Verbal Skill and Questionnaire Items .....	233
Attitude and Questionnaire Items .....	233
Sex and Questionnaire Items .....	234
Proficiency and Questionnaire Items .....	235
Holistic Score and Questionnaire Items .....	236
Prompt and Mode and Planning Time Questionnaire Items .....	237
Discussion .....	240

Tables .....	245
1. The Student Population Used in this Study by Grade Level, Sex, Subject and Section .....	245
2. Distribution of Verbal Skill Ratings .....	245
3. Distribution of Attitude Ratings .....	246
4. Distribution of Total Holistic Scores .....	246
5. Means and Standard Deviations of Proficient Students' and Less Proficient Students' Total Holistic Scores by Class, Section and Grade Level .....	247
6. Frequencies for the Different Modes of Writing Produced by Prompt A and Prompt B .....	247
7. Sex by Mode Frequencies .....	248
8. Proficiency and Mode Selection .....	248
9. Frequency of Student Responses to Questionnaire Items 3-25 by Prompt .....	249
10. Total Holistic Score Means, Standard Deviations and Numbers of Students by Prompt and Planning Time .....	249
11. Total Holistic Score Means and Standard Deviation by Prompt, Planning Time and Re-reading .....	250
Figures .....	251
1. A Schematic Diagram of the Differences in Planning and Writing Time for Versions A and B .....	251
References .....	252
Appendices .....	253
A. Writing Test Prompt Versions .....	254
B. Writing Questionnaire .....	255
C. Frequency of Responses for Questionnaire Items 3-25 .....	259

## Chapter 2

### Part II

#### COMPARISON AND ANALYSIS OF RATER RESPONSES TO THE ANCHOR PAPERS IN THE WRITING PROMPT VARIATION STUDY

Overview of the Study .....	260
Procedures of the Study .....	263
Raters as Subject .....	264
1. Student Raters .....	264
2. Novice Teacher-Raters .....	265
3. Expert Teacher-Raters .....	265
Anchor Papers .....	265
Training the Raters .....	266
Training the Student-Raters .....	266
Training the Novice Teacher-Raters .....	268
Training the Expert Teacher-Raters .....	268
Rating the Papers .....	270
Statistical Comparison of Results	
From Three Groups of Raters .....	272
Inter-Rater Reliability Estimates .....	272
Agreement Between Groups .....	273
Differences in Scoring Parameters	
Between Rater Groups .....	274
Differences in Score Distributions	
for three Rater Groups .....	275
Discussion of Results .....	276
Conclusion .....	281
Tables .....	282
1. Scoring Distribution for	
Each Group of Raters .....	282
2. Agreement Between Groups of Raters .....	283
2.1 Distributions of Scores Awarded	
by Three Groups of Raters .....	283
Figures .....	284
1. Facsimile of Grid for Period Two:	
32 Students .....	284
Appendix A .....	285

## Chapter 3

### BEYOND THE FIELD TEST OF WRITING PROMPTS: A PRELIMINARY INVESTIGATION OF THE NATURE OF RESPONSE VARIATION TO SELECTED FIELD-TESTED PROMPTS

#### Part I

THE EFFECTS OF ASSESSMENT PROMPT AND RESPONSE VARIABLES ON HOLISTIC SCORE: A PILOT STUDY AND VALIDATION OF AN ANALYSIS TECHNIQUE	
Introduction .....	317
The Data Base .....	318
The Analysis System .....	321
Response-based Variables .....	321
1. Legibility .....	321
2. Cohesion .....	322
3. Errors .....	322
Prompt-based Variables .....	324
1. Prompt-based Variables for Prompt 5 .....	324
2. Prompt-based Variables for Prompt 6 .....	325
Data Tabulation and Analysis .....	327
Results .....	328
Discussion .....	330
References .....	330
Tables .....	334
1. Means and Variances for Prompts 5 and 6 .....	334
2. Distribution of Papers, by Prompt and Score, Randomly Selected for Analysis .....	335
3. Legibility Ranking Scale .....	336
4. Analysis of Total Words, Legibility, Proportion of Total Chain Words, Proportion of Largest Chain Words and Proportion of Total Errors Across Prompts 5 and 6 .....	337
5. Across Prompt Analysis (Kendall's Correlation Co-efficient) .....	338

6. Score Related to Prompt-based Explicit, Elaborated and Inappropriate Responses by Prompt (Kendall's $\tau_c$ ) .....	339
7. T-Test Results, Means and Standard Deviations for Explicit, Elaborated and Inappropriate Responses Across Prompts .....	340
Figures .....	318
1. General Framework and Premise for the Study .....	318
2. Experimental Design .....	318

Chapter 3

Part II

A FIELD TEST OF EIGHT PROMPTS: SANTA CLARA  
COUNTY WRITING ASSESSMENT

The Scope of the Study .....	341
Developing the Prompts .....	341
Designing the Sample .....	346
Conducting the Writing Sample .....	348
Selecting Anchor Papers .....	350
Scoring the Papers .....	352
Results of the Study .....	355
Classroom Practice for Writing Competency Testing .....	376
References .....	379
Appendix A: .....	381
Members of the Santa Clara County Writing Committee Who Assisted in Developing the Prompts.....	381
Appendix B: .....	382
Sample Sizes for Testing Topic Performance .....	382
Appendix C: .....	384
Teachers Who Participated in Scoring Writing Sample Essays .....	384

Chapter 4

A STUDY OF THE CONSTRUCTION OF THE MEANINGS  
OF A WRITING PROMPT BY ITS AUTHORS, THE  
STUDENT WRITERS, AND THE RATERS

Part I

OVERVIEW OF GOALS AND PROCEDURES OF THE STUDY

Introduction .....	387
The Process of Prompt Construction .....	390
Target Students .....	390
The Assessment Day Procedures .....	391
Interview .....	391
Questionnaire .....	392
Interpretation of the Prompt by the Holistic Scorers .....	392
Pre-Scoring Interviews .....	392
Choosing Anchor Papers .....	392
Development of a Rubric .....	393
The Scoring .....	393
The Ranking of the Target Student Papers .....	394
Ranking Interview Procedures .....	395

Part II

THE WRITING PROMPT: DIFFERENCES IN EXPECTATIONS  
AMONG PARTICIPANTS IN A WRITING ASSESSMENT EPISODE

Introduction .....	397
The Prompt Development Session .....	398
The Holistic Scorers .....	399
Questionnaire Data on Student and Author/Rater Judgment of Task Demands .....	404
Differences Between Teacher and Student Interpretations of Task Demands .....	404
Questionnaire Responses and Total Holistic Score .....	406
Discussion .....	409

Chapter 4

Part III

THE WRITING PROMPT: THE PROCESSES OF INTERPRETATION  
AND PERFORMANCE THAT IT ELICITS IN STUDENT WRITERS

Assumptions Underlying the Analysis .....	411
Analysis of the Prompt .....	411
Cohesion and Contextualization .....	412
Cohesion and Unifying Purpose .....	413
Analysis of Oral and Written Responses to the Prompt .....	418
Nature and Relevance of Idea Units .....	418
Procedure for Analyzing Oral Responses .....	419
Procedure for Segmenting Written Compositions .....	422
A Typology of Idea Unit Functions .....	424
Interactional .....	425
Think-Aloud Process .....	425
Planning for Composing .....	426
Past Experience .....	426
Reference to the Prompt .....	427
Exact Wording .....	427
Paraphrase .....	428
Reference to the Prompt as a Whole .....	428
Typical Pattern: Reference to Prompt/Comment .....	428
Reference to Shared Knowledge .....	429
Basic Reading Ability Factors .....	432
Evaluative and Narrative Functions .....	433
Evaluative Strategies .....	436
Attribution .....	436
Generalized Attribution .....	438
Generalization .....	440
Comparators .....	442
Comparison of Oral and Written Units .....	445
Discussion .....	450
Tables .....	401
1. Percentage of Responses by Students and Group of Evaluators of Selected Questionnaire Items .....	401

2. Kendall's Correlation Coefficients of Responses to Selected Questionnaire Items with Total Holistic Score .....	408
Figures .....	388
1. Diagram of Assessment Episode .....	388
2. Rubric .....	393
3. Revised Rubric .....	394
4. Criteria for Segmenting Written Discourse .....	421
5. Transcription Symbols .....	422

Chapter 5

HOLISTIC ASSESSMENT AS LONGITUDINAL  
MEASURE OF STUDENT GROWTH:  
INTERPRETING STUDENT TASK CONSTRUCTIONS

Part I

UNEXPECTED DIRECTIONS OF CHANGE IN  
STUDENT WRITING PERFORMANCE

Introduction .....	473
Developmental Unevenness and Current Measures .....	475
The Ill-Defined Writing Task .....	477
The Drake Longitudinal Sample .....	480
A Discourse Typology for Developing Writers' Texts .....	482
Choosing a Strategy .....	485
References .....	501
Tables .....	503
1. Drake Writing Assessment Topics .....	503
2. Choosing a Strategy .....	504
Figures .....	505
1. Sample Population .....	505
2. A Discourse Typology .....	506
3. Proposition Structure of Three Discourse Strategies .....	507
4. Hypothesized Chronology of the Acquisition and Development of Writing Strategies .....	508

Chapter 5

Part II

BEYOND HOLISTIC SCORING: RHETORICAL FLAWS  
THAT SIGNAL ADVANCE IN DEVELOPING WRITERS

Introduction .....	509
Purpose of the Study .....	510
The Data Base .....	513
The Student Writers .....	516
The School Writing Program .....	516
The Impressionistic Quality Rating .....	516
Selection of Cases for Analysis .....	517
The Typology .....	518
Rhetorical Function .....	520
Discourse Strategy .....	520
Proposition Structure .....	520
Primary types .....	522
Application of the Typology .....	524
Problems in Application of the Typology .....	527
Results of the Hypotheses .....	531
Topic/Year 1973 "Event" Topic .....	532
Topic/Year 1974 "Be a non-human thing" Topic.....	533
Topic/Year 1975 "Be another Person" Topic .....	533
Topic/Year 1976 "Invention" Topic .....	534
Topic/Year 1977 "Space object" Topic .....	534
Topic/Year 1978 "Influence" Topic .....	535
References .....	537
Tables .....	539
1. Drake Writing Assessment Topics .....	539
2. New Score Means for Overall Sample and Individual Cohorts .....	540
3. New Score Means for Each Grade Level Compared to Overall Sample Means for Each Topic/Year.....	541
4. Topic Years According to Cohorts .....	542

PROPERTIES OF WRITING TASKS:  
A STUDY OF ALTERNATIVE PROCEDURES FOR  
HOLISTIC WRITING ASSESSMENT

EXECUTIVE SUMMARY

Leo Ruth, Project Director

Background and Rationale

Evaluation is an essential step in the process of instruction and in determining the accomplishments of school programs. The evaluation model being refined in the Bay Area Writing Project's study of writing assessment is the so-called "holistic" approach to appraisal of the writing samples produced in response to given prompts. Holistic assessments test writing by calling for a student to produce an actual sample of writing, which later is read and rated according to systematic procedures variously derived. The rapid growth of this style of writing assessment has, however, outstripped knowledge of how components of the model interact. Needed is a comprehensive study of interactions between writing task, student response, and reader's rating in the course of making a holistic writing assessment.

Often the first question asked when evaluators embark on a holistic assessment is "What topic should we use and how should we word it?" The writing task, whatever its nature, serves as the prompt or the stimulus for the writing performance evoked. Inasmuch as the prompt stands in dynamic relation to the piece of writing it initiates, evaluators presumably would desire to know how particular properties of prompts might affect the performance of writers in both intended and unintended ways. Yet surveys of relevant literature reveal surprisingly little theory or data regarding the properties of writing prompts.

Given the long history of the use of written examinations in this country, one might expect to find a body of knowledge drawn from theory, research, and practice which guides the design of the prompts (topics, tasks, assignments, exercises, questions, instructions, visual stimuli) used to elicit specimens of writing for purposes of evaluation. Yet, the search of the literature presented in Chapter 1 reveals that relatively few details are yet available from educational or psychometric research against which one might check current practice in designing writing assessment prompts. Various researchers in the field of writing have noted this particular gap in the available research on writing assessment (Odell,

Cooper, and Courts, 1978; Odell, 1979; Greenberg, 1981; Hoetker, 1981).

Psychometrics, for example, has dealt with a restricted range of concerns in relation to developing writing prompts. Textbooks on measurement commonly deal with test construction, but their focus tends to be more on the technology and methodology of measurement than on the analysis of substantive properties of the essay questions that comprise writing tests. Psychometrics provides methods of applying statistical approaches to the appraisal of the "behavior" of writing prompts or objective test items during field trials and actual test administrations. But psychometrics does not offer guidance beyond the most general sorts of principles for actually designing and wording the writing tasks.

Through the agency of a "field trial," psychometrics provides a self-evaluative procedure for eliminating faulty essay questions or other test items, but psychometrics does not have empirically derived analytic methods for determining what has gone wrong in framing the essay question (Payne and McMorris, 1967). The essay questions that survive the field trial are presumed to be "meaningful" to the larger population to be tested. Though psychometrics strives to be scientific, the procedures used to construct and validate essay questions are essentially subjective and scientifically crude in the sense that there is no theory of information processing or of cognition underlying the design of writing tasks.

Standardized testing procedures pre-suppose that the examiner and the student writer share a common language and uniform interpretations of tasks. Yet, Mellon, for example, in his discussion of the first National Assessment of Educational Progress Writing Assessment commented on the difficulties of the NAEP assessors in formulating writing topics (even with the help of experienced personnel from the Educational Testing Service). Mellon wondered why a seemingly straightforward topic like "Going to School" produced, on the part of some writers, such mundane accounts of scenes along the way, while it elicited descriptions of fantasy occurrences in others. This is only one example selected from many that clearly indicates particular topics do not guarantee the activation of particular levels of thought and forms of discourse. The function of a writing task as a stimulus event cannot be understood until the examiner determines how constraints of language use and interpretation are operating in the writing assessment situation.

Despite the fact that researchers in educational measurement have acknowledged question-wording effects (French, 1966; Coffman, 1971), the larger theoretical issues of essay task wording and typology have seldom been addressed. There has been little research to determine the frequency, magnitude, or underlying nature of essay task-wording effects. Wording effects have tended to be treated anecdotally or incidentally

in the context of other objectives of writing measurement study. (For example, see anecdotal treatments of wording effects in Godshalk, Swineford, and Coffman, 1966; Winters, 1980; Smith, 1980; Lloyd-Jones, 1977; Michigan Survey Research Center Interviewer Manual, 1969, etc.) Current procedures for "item analysis" do not lead to "analysis" of the linguistic properties inside the texts of essay questions which might affect their interpretations by the persons responding to them. Sax acknowledges that the procedures called "item analysis" are such that "unfortunately [they] cannot reveal if ambiguity results from the students' lack of knowledge or from poorly written items" (1974, p. 233).

There are, however, two other sources that promise useful information to guide the development of writing prompts. These sources are discourse theory and practitioner knowledge. Discourse theory, which deals with the development of full texts in speech and writing (Kinneavy's definition, 1971), does provide a source of knowledge that influences the design of writing assignments for both instructional and testing occasions. Discourse theory, with roots reaching back to antiquity, attempts to describe the varieties of discourse and the elements which contribute to the making of texts, spoken and written. Various schools of thought have emphasized one or another element such as the writer, the subject, or the audience, either individually or taken together. A fourth element, function (also called purpose or aim) arises from the interaction of writer, subject, and audience. Related to all of these is the concept of form (also called mode or genre). The basic problem in dealing with this field is the very richness of the area and the problems attendant upon selecting from among competing theories. These various discourse theories provide the categories of forms and functions of writing that frequently turn up as specifications and constraints in writing prompts.

Practitioner knowledge tends to be undervalued: It has neither the status of theory nor of empirically tested knowledge. Yet practitioner knowledge develops through a continuing involvement in the unique and particular experiences of the profession. This practical knowledge may denote a high level of competence, but it remains essentially an individual theory of practice so long as it remains private. However, some of the privately-held theories of practice eventually "go public" through textbooks, curricula, and other media. Once public, these works are subject to critical processes which may lead to their verification, validation, and refinement. Thus, in time, there develops a form of theory that might be called "a theory of idealized practice." This source offers knowledge to use in developing normative "rules" that may be recommended on the basis of successful practice for designing prompts. The practical wisdom available from the experience of professionals in topic writing needs to be collected, sorted, appraised, and used to formulate propositions to guide the

design of writing test prompts.

There are indications in the literature that researchers, evaluators, and teachers are not sufficiently aware of the nature and range of task effect upon writer performance. Bereiter, Scardamalia, and Bracewell have noted:

Developmental research in writing has generally proceeded by assuming that the task was construed the same by all subjects and therefrom inferring differences in competence. But this is quite an inadequate way of going after any deep understanding of cognitive development in writing (1979, p. 5, as quoted in Polin, 1980, p. 8).

Local school assessors often provide an all-purpose essay topic and generalize broadly from the student's response to this single measure to predict performance on any writing occasion, assuming that "good writing" is a unitary construct, not subject to differences in topic and purpose. Similarly, some schools offer topics intended to stimulate exposition or argument but which instead measure the writer's general knowledge rather than his/her capacity to mount an argument or sustain a piece of exposition.

In addition, many testers insist that "audience should be specified," but they fail to make distinctions about what kinds of audience. A California writing assessment once specified audience as follows: "Write a letter to your pen-pal." But a study by Marion Crowhurst (RTE, 1978) found that students produce more sophisticated and complex syntax when writing for a teacher or general reader than when writing for a peer. If testers want to evaluate maturity of syntax, then a prompt which specifies an unsophisticated audience is counterproductive.

The writing stimulus, whatever its nature, is a springboard. Researchers and assessors must not lose sight of the linguistic, cognitive, and social reverberations it may set off in the writer. Yet, as far as we can determine, there are a number of interactions between writing stimulus and writers that have not yet been explored. At issue is the nature of the interaction between the topic and any accompanying instructions (written or verbal) and what these may "prompt" a writer to produce. Since it is known that choice of topic with instructions can affect the performance of the writer, how do we know whether the prompt given is actually causing the student to perform either typically or as well as he actually is able to perform in the areas we would like to assess?

There are, then, potential sources of invalidity in the conduct of holistic assessment at the point of interpreting the writing task and in the reading of the writing samples produced. There is a potential for possible miscommunication in the writer's interpretation of and response to the writing

problem set in a holistic writing assessment. These are the issues that have been addressed in a series of inter-related research studies presented in the following report. The several studies provide data, theory, and recommendations for practice that will enable the authors of writing assessment prompts to understand more exactly the nature of cues and constraints within the prompts and how these elements affect the performance of students writing the papers and the judgments of evaluators reading the papers. The data and recommendations from these studies are expected to have significant implications for practices in the generating of prompts intended to initiate the display of competence in writing assessments.

### The Research Problem

The principal research purpose governing the work of all dimensions of this NIE/BAWP Writing Assessment Project is to increase our knowledge of the properties of writing test prompts and the compositions they initiate in holistic writing assessments. There is considerable evidence to enable us to conclude that particular prompts do not perform as uniform stimuli across populations on various testing occasions to activate expected levels of thought and forms of discourse. This investigation asks the questions:

1. What elements in the content and structure of writing test prompts influence the production of compositions during writing assessments?
2. What makes a prompt "good" for eliciting writing samples to be rated under conditions of "holistic scoring" procedures?
3. How are "good" writing prompts generated and tested systematically?
4. What changes occur in maturing students' responses to writing prompts in successive annual assessments, and what do these changes reveal about how we should be assessing developing abilities in writing?
5. What sources of knowledge do we have to guide current practice in designing effective writing prompts?

The several chapters of this report address these basic research questions.

## Description of the Research

Theoretical Grounding of the Research. Writing research has only recently emerged as a distinct field of inquiry. As an interdisciplinary field it draws from branches of linguistics, psychology, anthropology, education, cognitive science, social science, and literary study. It would be difficult to trace in this summary of our report all the crossings of discipline boundaries undertaken within each of the five studies presented here. But the particular significance of one theoretical field is clear: we noted in our research proposal that the act of writing begins with an act of reading comprehension as the writer encounters the text of a prompt. Given our observation of the variability of interpretations of writing tasks, we were drawn to the constructivist theories of reading which we began to adapt into an evolving constructivist model of writing task comprehension.

Our conception of the nature of reading comprehension has been changing within the last few years as a result of theoretical and experimental work in fields such as psycholinguistics, cognitive psychology, artificial intelligence and other disciplines. An increasing number of studies in these fields have given attention to comprehension and memory, both considered as constructive processes. The "constructivist" position considers comprehension to be a complex process which depends on world knowledge, and intra-sentence content, as well as inter-sentence context. The constructive process feeds on the interaction of information presented through text structures, the context of that information, and the existing knowledge framework of the reader. Reading, in effect, is a specific application of general cognitive processing skills and strategies.

Our study of properties of prompts is based upon analyses which propose a constructivist theory of writing task interpretation. Writers choose rather freely among cues embedded in the prompt, both honoring and ignoring elements which may enable the writer, with varying degrees of success, to match the test-maker's intention and expectation. The meaning potential of a given prompt is thus relative to the linguistic, cognitive, and social reverberations set off in the writer. This theory draws on schema-based theories of information processing and socially-based theories of communicative competence. This theory of the nature of the prompt shows how particular properties of topics--ideational, interpersonal, and textual aspects--function simultaneously in the assessment episode in shaping the written responses that are formed to meet the perceived requirements of the topic.

The constructivist theory of task meaning being developed here has significance for generating prompts for holistic writing assessments. The problem is not one of simply forming "better" prompts but of understanding how a host of properties

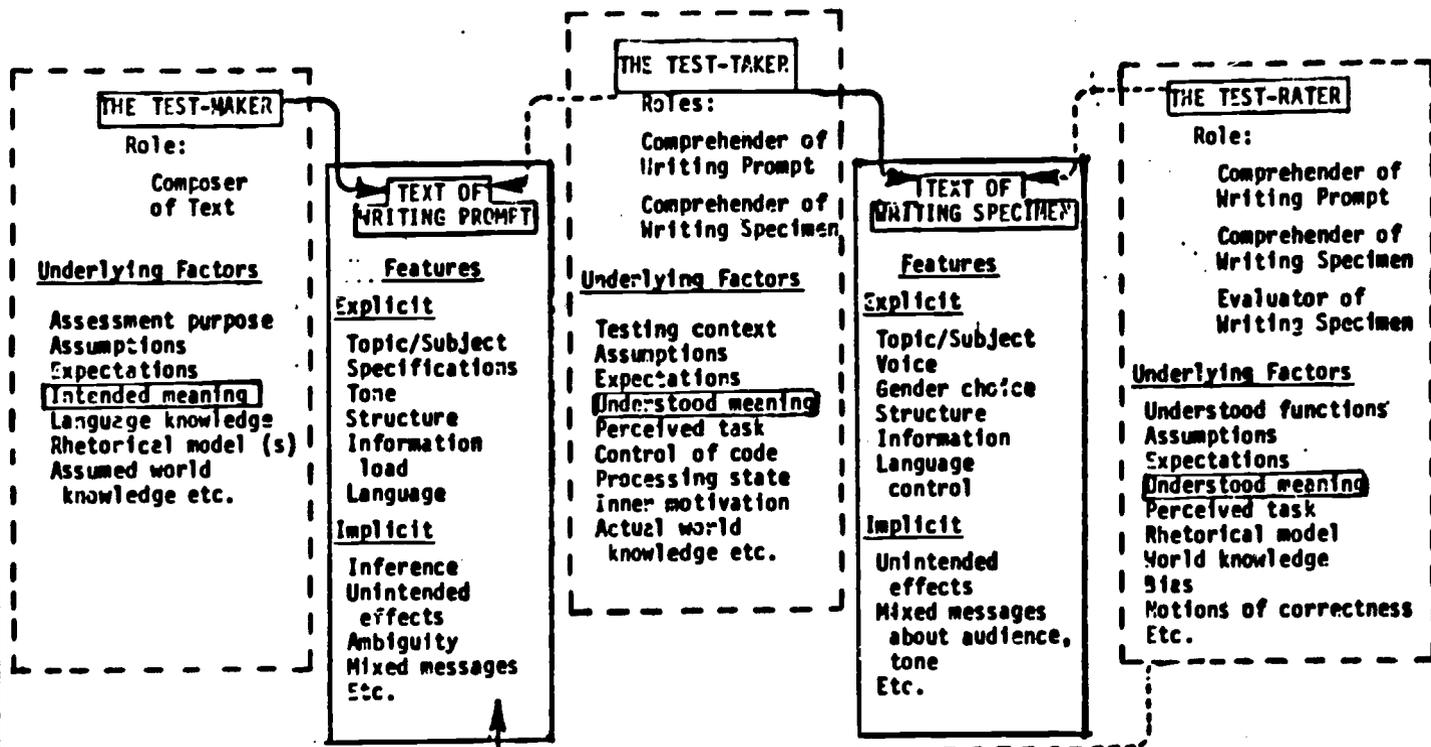
may influence what a writer "makes" of any prompt given to him to initiate his display of writing competence: the same prompt put to two different people may not be perceived as the same prompt. The properties of tasks and what they elicit in writing performances are contingent upon the language of the prompt, the world knowledge of the participating writers, and the social roles of the participants in the "speech event" of a writing episode.

The Model of the Writing Assessment Episode. In thinking through our strategies for approaching our research questions, we found it useful to keep in mind the simple visual model of an assessment episode as depicted below. There are varying degrees of abstraction and formalization in the construction of models. Some achieve the strength of predictive models of explanations of scientific phenomena. We created ours, however, as a scheme for consolidating and ordering the concepts we wanted to keep in mind. For a model makes knowledge accessible for display, manipulation, and interpretation. As a manageable representation of entities in the real world that it represents, a model becomes a useful guide for making interpretations and decisions about the real world it represents. In this case, we wanted to keep in mind the unity of the assessment event and the interactive relationships of the participants, processes, and products that comprise it. So, early in the project we began with a much simplified version of the model now presented here. This version is still labeled "tentative," indicating that it invites further modification. Essentially, there are three basic processes depicted in our model that we sought to understand (see following page).

1. THE PROCESS OF DESIGNING THE PROMPT. What goes on in the test-maker's mind during the construction of a prompt? What wording choices does the tester make to convey accurately his/her intentions? How does the writer of the prompt make decisions relating to rhetorical specifications of mode, role, audience, purpose? What assumption about world knowledge and cognitive developmental level of the writer does the test-maker hold?
2. THE PROCESS OF PERFORMING/COMPOSING. What does the student writer understand his/her task to be? Which of the tester's cues and constraints does the student writer recognize, understand, and use? Given the writer's interpretation of the task, how does he/she go about accomplishing it?
3. THE PROCESS OF RATING. What does the evaluator understand the intention of the prompt to be? How does the evaluator interpret the intention of the writer in relation to the evaluator's own

AN EVOLVING MODEL FOR STUDYING THE WRITING ASSESSMENT EPISODE:

PARTICIPANTS → PROCESSES → PRODUCTS



----- = Cognitive Processes  
 \_\_\_\_\_ = Visible Product

- The basic categories of the model were proposed by the project director, Leo Ruth, at a staff meeting on 10/22, 1980. Many refinements of the current version reflect the contributions of the participants in our weekly conceptual brainstorming meetings, attended in the first year of the project (1980-81) by Sandra Murphy, Catherine Keech, Paul Ammon, Judith Langer, Donald Leu, Charles Kinzer, and Karen Carroll, and other visitors.

perception of the task? How do the evaluators' biases--preferred discourse model, notion of the relevance or irrelevance of lower order scribal skills, preferences for displays of creativity, originality, reasoning or the lack of these, opinions and knowledge of the world--enter into the assigning of quality ratings?

### The Principal Studies

The principal studies of the NIE funded Bay Area Writing Project Writing Assessment Project are listed below:

1. Designing prompts for holistic writing assessments: Knowledge from theory, research and practice, by Leo Ruth and Catharine Keech
2. Effects of variation in a writing test prompt upon holistic score and other factors, by Catharine Keech, Don Leu, Sandra Murphy, and Charles Kinzer
3. Beyond the field test of writing prompts: a preliminary investigation of the nature of response variation to selected field-tested prompts, by Charles Kinzer and Sandra Murphy
4. A study of the construction of the meanings of a writing prompt by its authors, the student writers, and the raters by Sandra Murphy, Karen Carroll, Charles Kinzer, and Ann Robyns
5. Holistic assessment as longitudinal measure of student growth: Interpreting student task constructions, by Catharine Keech and Elissa Warantz

## HIGHLIGHTS OF THE STUDIES

### PROPERTIES OF WRITING TASKS IN HOLISTIC WRITING ASSESSMENTS

#### STUDY 1. DESIGNING PROMPTS FOR HOLISTIC WRITING ASSESSMENTS: KNOWLEDGE FROM THEORY, RESEARCH AND PRACTICE

##### Part I: Sources of Knowledge for Designing Writing Test Prompts

Leo Ruth

#### A. The Problem

1. What sources of knowledge do we have to guide current practice in designing effective writing prompts:
2. What makes a prompt "good" for eliciting writing samples to be rated under conditions of "holistic scoring?"
3. How are "good" writing prompts generated and tested systematically?

#### B. Outcomes

1. Five principal sources of knowledge were studied:
  - a. Psychometrics
  - b. Professional wisdom
  - c. Discourse theory
  - d. Writing research
  - e. Selected Practices
2. The selected practices were reviewed in draft form by a panel of teachers, researchers, and professional test-makers.
3. The report on practice, revised in the light of critic-reviewers' recommendations appears as Part II of this study.
4. Ideas from the other four principal sources of knowledge about writing test prompts are summarized in four guides beginning at the end of Part I, Chap. 1.
  - a. The Philosophy of the Writing Assignment for Assessment
  - b. Guidelines for Developing Prompts for Writing Assessment
  - c. A Tentative Schema of Writing Tasks for the Development of Assessment Prompts
  - d. Schema for Task Analysis of Prompts

#### C. Issues and Recommendations

1. The manipulation of variables is problematic in experimental research. The notion of the introduction of a "variable" into a prompt presupposes that it is perceived as something significant.

2. There is currently no well-established rational, rule-governed procedure for generation of writing test prompts. We need to discover and test systematic and theoretically sound rules which enable us to design appropriate prompts for given test-writer-task occasions. But we need to be wary of going so far that we turn the prompt development process into a "paint-by-numbers" approach.
3. There are a number of text characteristics that might be taken into account:
  - a. Text length
  - b. Density of information
  - c. Density of new information
  - d. Density of arguments in propositions
  - e. Capacity to inspire interest
  - f. Cohesive features
  - g. Presence or absence of information
4. Prompts need also to be examined for the cognitive demands they make:
  - a. Language production requirements
  - b. Presupposed genre knowledge
  - c. Memory and retrieval requirements
  - d. Reasoning requirements
  - e. Explicitness of task demands and potential for recognition
  - f. Developmental factors in interpretation
  - g. Motivational potential
  - h. Presupposed world knowledge
5. Field trials are essential, but new steps need to be taken. Instead of merely developing statistical indexes of difficulty, we need to go beyond the trial to conduct task analyses and find out why writers responded the way they did.
6. Raters and evaluators need to recognize a range of allowable variation in the writer's interpretation and response to the prompt.
7. Implications for teaching include:
  - a. Teaching students how to "read" a writing test prompt:
    - how to separate out task components
    - how to read for intention, implication
  - b. Teaching students strategies for managing constraints

- c. Understanding the distinctions between "real" writing, "school" writing, and "test" writing
- d. Understanding that the process of writing under time constraints in a test situation is different from natural processes of writing.
- e. Teaching students how to achieve flexibility in the range of roles they can assume and audiences they can address.

Part II: Practices in Designing Writing Test  
Prompts: Analysis and Recommendations

Catharine Keech

A. The Problem

1. How should students be asked to write and what should they be asked to write about if we wish to measure their real writing abilities?
2. What basic principles guide professional test-makers in answering the first question?

B. Outcomes

1. The entire report constitutes the outcome of the study. What follow are only a few of the interesting findings.

Selection of the Subject (and Problems of Prior Knowledge)

- a. Topics that encourage introspective or autobiographical writing by drawing on personal experience are both motivational and intentional.
- b. "Choosing interesting topics and expressing them simply will ensure that most students can find something meaningful to say in their writing."
- c. Content specifications can "cue" students to see the full potential of the topic, explain the writer's options, and/or clarify the tester's expectations. However, there are two dangers of elaborate cuing:
  1. Poor readers may be put off by the sheer amount of text; and
  2. Students could mistake the cuing as "commands" rather than as "suggestions."
- d. Hypothetical situations (asking students to imagine, pretend, or wish) make good subjects for writing; however, the need to use the conditional or subjunctive verb phrases can present problems for some students.
- e. Problems relating to differences in prior knowledge among students can be avoided by providing a body of relevant texts or a literary passage to be commented on or argued from.

- f. Controversial issues can provide especially appropriate topics for tests of persuasive writing. But "controversial topics can create special problems for holistic assessors if the subject arouses strong biases in both student-writers and teacher-readers" because readers can be distracted from the task of evaluating writing ability by the nature of the written content.

### C. Issues and Recommendations

The report is filled with many recommendations and discusses many issues. Only a small sample of these is provided here.

1. The "describe trap" can be avoided by substituting the instruction, "Tell about."
2. Instructions are useful, but too many--i.e., overprompting--can become constraints.
3. A very helpful constraint in instructions is a suggestion to focus on one aspect of a larger topic.
4. The "special conventions trap" can result in student misunderstanding of the intent of the assessment--i.e., is knowledge of the particular convention being tested as well as the overall ability to write--and too much attention gets paid to the formal aspects (e.g., the business letter) rather than to the production of a good piece of writing.
5. The "pick-a-form trap" may seem free, but teacher-raters often have preferred forms that get rewarded, while other forms are penalized.
6. Specifying number of works, sentences, paragraphs is an unnecessary--and usually misunderstood--constraint that does not qualify as helpful instructions. Such specification inhibits fluency, and emphasizes minimum rather than optimal performance.
7. Specifying audience may be useful in some testing situations, but can cause problems in others. Merely mentioning an audience is not sufficient. "When audience is specified in a writing task that provides a full rhetorical context and a clear purpose for writing, the effect is to produce a more realistic writing problem."
8. Avoid hidden biases and loaded words and phrases. Students sensing the bias may write from their conception of what the test-maker believes, rather than from their own perceptions and opinions.

STUDY 2.        EFFECTS OF VARIATION IN A WRITING TEST PROMPT  
                 UPON HOLISTIC SCORE AND OTHER FACTORS

Part I:    Effects of Two Versions of a Writing  
            Prompt Upon Holistic Score and Writing  
            Processes

Don Leu, Catharine Keech, Sandra Murphy,  
and Charles Kinzer

A.    The Problem

1.    The purpose of this study was "to investigate effects of variation in the specification of audience on a timed writing test..." The study involved 114 high school students, grades 10-12, from one teacher's four classes.
2.    The day after students wrote in response to one of the two versions of the prompt, they completed a questionnaire intended "to elicit information about their attitudes toward the prompt, their recollections of the writing process during the timed writing tests, and their descriptions of their other writing behaviors and attitudes, both in and out of school."
3.    Two questions were of particular interest in examining the data drawn from the writing test and the questionnaire:
  - a)    Which variables are associated with high holistic scores?
  - b)    Which variables are associated with differences in the prompt?

B.    Outcomes

1.    Differences in the two versions of the prompt did not produce a difference in the performance, as measured by total holistic score means. But does this mean that providing additional information about the writer's audience and rhetorical purpose does not affect the quality of performance in a timed writing test?
2.    Students' choice of writing mode was affected by the different prompt versions. Students writing on Version B (with a specified audience) produced 20 alternate mode papers (16 letters, 3 journal entries, 1 essay); students writing on Version A (with an unspecified audience) produced only seven alternate mode papers (5 letters, 2 journal entries).

3. More planning time was spent on the prompt version with the less specific audience definition.
4. Nearly three-fourths of the students reported that they would have written very differently if they had had more time. This suggests that student performance on a short, timed writing test may differ significantly from performance on a regular classroom writing assignment.

### C. Issues and Recommendations

1. "Although one task constraint, degree of audience specification, was manipulated experimentally, the study as a whole should be viewed as descriptive and explanatory rather than as an exactly controlled experiment. Viewed in this light the study offers a number of provocative leads for needed investigation."
2. Additional research is necessary before concluding that specification of audience and rhetorical purpose does not affect the quality of a student's performance on a timed writing test. "Clearly what is needed is additional research on pairs of prompts in which the difference in audience specification is made more evident. The failure to find differences in the quality of writing between these two versions may have been due to the fact that they both represent middle range examples on the continuum of audience specification."
3. The failure to find topic-related differences in the quality of writing may have been related to the fact that this was only a single writing assessment episode, and, as such, provided no opportunities to gain insights into whether this particular performance was truly representative of these students' usual writing practices. Observing the same students over several writing tasks would undoubtedly reveal patterns--such as "whether some kinds of tasks consistently require more planning time than others...and whether planning consistently produced better products."
4. Whether degree of audience specification can serve to trigger more and/or faster mental associations or responses in the student-writer has important implications for evaluation of instructional programs and for research seeking to assess performance differences among able as well as less able student writers.
5. What are the effects of time constraints on quality of writing, on forms selected, on amount of planning time, amount of revision, etc.?

Part II: Comparison and Analysis of Rater  
Response to the Anchor Papers in  
the Writing Prompt Variation Study

Catharine Keech and Mary Ellen McNelly

A. Problem

1. This part of the study represents an attempt to "identify factors in writing test design which may contribute to the instability of student performance on direct measures of writing" by comparing "student and teacher evaluative criteria in scoring student essays." Two focuses of study:
  - a) An examination of task interpretations of student writers, revealed through their discussion of the qualities of the anchor papers during a training session that prepared them for a holistic scoring of their own papers;
  - b) Comparisons of differences between students and expert raters in terms of interpretations and evaluative criteria.
2. This part of the study also included an attempt to explain the apparent lack of congruence in task interpretation between students and testers (or teachers) which leads to apparent failure in student performance, as determined by teacher-raters. The problem is approached through two primary questions:
  - a) Do writers and raters have the same perception of what the task is?
  - b) Do writers and raters have the same perception of what constitutes a successful completion of the task?

B. Outcomes

1. A major finding was "that some students do in fact interpret the writing task in ways which differ from teacher-rater task interpretation."
2. Differences between the three groups (student raters, novice teacher raters, expert teacher raters) in the inter-rater reliability could be accounted for by examining the nature of and attitudes toward the training for holistic scoring and the initial defining of scoring categories.

3. Versimilitude on writing assessment tests may actually be penalized by teacher-raters because such "real writing" (e.g., a letter) is, in fact, different from "real English themes" (or "school writing"). "The conflicting frames of reference provided by the simulated life-like communicative situation and the actual writing context--a classroom test occasion--can be successfully integrated only by the most able student."

C. Issues and Recommendations

1. One difficulty encountered was that the process of choosing anchor papers to define a scoring range--a critical step in any holistic assessment--requires the juggling of sometimes contradictory constraints.
2. Lack of agreement among the three groups of raters could be explained, in part, by an examination of the values and criteria they expressed, with regard to particular anchor papers, during the training sessions.
3. An issue raised which requires further study is the differences between responses to typed versions of papers (used in reading and rating anchors during the training session) and original handwritten versions of papers used during the actual reading of the test papers.

STUDY 3. BEYOND THE FIELD TEST OF WRITING PROMPTS:  
A PRELIMINARY INVESTIGATION OF THE NATURE  
OF RESPONSE VARIATION TO SELECTED FIELD-  
TESTED PROMPTS

Part I: The Effects of Assessment Prompt  
and Response Variables on Holistic  
Score: A Pilot Study and Validation  
of an Analysis Technique

Charles Kinzer and Sandra Murphy

A. The Problem

1. In examining writing task variables in holistic writing assessments, the focus in this part of the study is on possible relationships between assessment prompts, written products, and holistic scores. The study represents a preliminary effort to determine ways to identify unintended effects of prompt and response variables in relation to holistic scores.
2. Two major goals of this study:
  - a) To investigate prompt effects and response variables which might result in discrepancies in holistic scores.
  - b) To validate an analysis system for looking within specific prompts.

B. Outcomes

1. The results of the study support the viewpoint that "effects of writing prompts are related in student responses and influence holistic scores."
2. The study found that "task demands can be related to score, and that an analysis system modeled on the one employed herein can be valuable in determining how well students respond to a prompt, and in determining how difficult a prompt is."

C. Issues and Recommendations

"The refinement of procedures for the identification and analysis of effects of writing prompts is a necessary step toward devising ways to help teachers and evaluators improve the formulation of prompts for writing assessment and instruction. Given the widespread use of direct measures for evaluating writing competence, every effort should be made to control for unintended effects in the prompts used in these assessments."

Part II: A Field Test of Eight Prompts: Santa Clara County Writing Assessment

Gerald Camp

A. The Problem

The goal of this part of the study was to obtain data on the performance on each of the eight prompts--developed for use in the Santa Clara County Writing Assessment--in typical county schools identified as low, mid-range, and high based on the performance of students on other measures of writing ability.

B. Outcomes

1. Teacher-readers were asked questions after completing the scoring of the papers. The first two questions asked whether the teachers felt that any of the prompts produced noticeably weaker or noticeably stronger writing than others. The responses to these questions showed very little agreement among the teacher-readers as to which of the eight prompts were less and which were more successful.
2. Almost all of the prompts were successful in producing the full range of writing quality desirable in such a sampling.

C. Issues and Recommendations

"Writing short timed essays on subjects students have had no opportunity to prepare for requires somewhat different skills from those needed for the usual school writing assignments. It seems only reasonable, therefore, that students should be given some training in these skills before being required to demonstrate their competency with them.

The best training for tests of this sort is frequent in-class practice with the kinds of topics which will be used for actual testing. Students can learn how their test papers are judged if they also learn to score each other's papers themselves using the holistic method."

STUDY 4: A STUDY OF THE CONSTRUCTION OF THE MEANINGS OF A WRITING PROMPT BY ITS AUTHORS, THE STUDENT WRITERS, AND THE RATERS

Part I: Overview of Goals and Procedures of the Study

Karen Carroll and Sandra Murphy

Part II: The Writing Prompt: Differences in Expectations Among Participants in a Writing Test Episode

Charles Kinzer, Karen Carroll and Sandra Murphy

Part III: The Writing Prompt: The Process of Interpretation and Performance That It Elicits in Student Writers

Sandra Murphy, Karen Carroll, and Ann Robyns

A. The Problem

1. The problem of student misunderstanding of teacher intention in comprehending the writing assessment is more grave and lasting in large-scale writing assessment episodes, in which opportunities for feedback do not exist once the writing has begun, than in a classroom setting in which the teacher is accessible.
2. Investigation of the student interpretation of prompts involved interviewers recording the interpretations of student-writers prior to writing and immediately afterwards in the assessment episode.
3. Investigating "teacher intention" involved asking teachers, while in the process of constructing the prompt to be used in the assessment, to verbalize their expectations of students' responses. Later other teachers predicted student interpretations and prompt-based outcomes.

B. Outcomes

1. Data in the study indicated that in this assessment, in spite of careful construction of the prompt, expectations of evaluators differed from those of student-writers; in addition, the data indicated that implicit task demands may have an effect on the way students' papers are evaluated and ranked.
2. "Naive scorers' prompt-based expectations were extremely similar to those expressed by the authors

of the prompts."

3. "Teachers and students often perceived task demands differently. One major difference was that the prompt authors lent weight to all parts of the prompt, while some of the student-writers focused on only a few selected aspects of the prompt."
4. "In the larger group of students, the results indicated that students who rated the prompt as easy to read and understand tended to receive higher holistic scores. High scores were also associated with agreement that it was important to explain why the subject chosen was a problem, and that it was important to suggest a solution to the problem. The results also indicated that students who agreed with the statement that it was important to explain how the solution/correction of the subject chosen would improve Central High tended to receive higher scores. Though the measures of associations are low, they do support the view of the importance of the effects of writing task comprehension in a writing test solution."
5. "An underlying assumption was that the student accesses information in the process of reading the prompt which he or she then uses in composing a written response. The construction of an interpretation of the prompt is thus viewed as an integral and important part of the composing process. In the process of interpreting the prompt, the student may selectively focus on some segments, ignore others, and interpret the ones assimilated differently than do the evaluators in the assessment. Thus, the student may "construct" an interpretation of the prompt which differs from the interpretation of other participants in the assessment."
6. "In order to determine how students used their initial interpretations in their written compositions, segments of the oral interviews were compared to segments of the written compositions. Both similarities and differences were found between the oral language of the interviews, and the written language of the students' compositions. Basically the similarities were found in content, while the differences were found in rhetorical function."
7. "The study suggested that context plays an important role in both the reading and the writing process. However, the issue was not simply one of "knowledge" or lack of it; rather, the issue was whether or not a context should be

STUDY 5: HOLISTIC ASSESSMENT AS LONGITUDINAL MEASURE  
OF STUDENT GROWTH: INTERPRETING STUDENT  
TASK CONSTRUCTIONS

Part I: The Theoretical Model

Catharine Keech

A. The Problem

1. "Why do performance scores on writing tests so often fail to improve in neat positive intervals for individuals and groups?"
2. One aspect of performance variation is the phenomenon of apparent regression--the U-shaped learning curve--during the acquisition of new skills and performance strategies.
3. Another aspect of performance variation is "the ill-defined nature of the writing task, which allows the student writer to determine, to a large extent, the difficulty level of what he attempts regardless of the tester's intention."

B. Outcomes

1. "Apparent regression" may simply be one manifestation of the non-linear nature of improvement in developing writers. Errors committed by writers-in-transition (i.e., developing writers who are in the process of learning a more advanced writing strategy) don't necessarily indicate that they are less able performers than before the new learning began. Rather, these errors, when seen in context, may be signalling movement toward a higher level of competency. That is, writers-in-transition can be perceived as practicing their newly-learned writing strategies.
2. In some writing assessments (or writing assignments) the complexity of the cognitive task demanded may distract the writer's attention away from rhetorical features--even those that the writer has previously mastered.
3. "How the student defines the rhetorical task at the text or discourse level," how--and whether--the student performs whole-text planning (that is, conceptualizing an "artifact" that he or she is about to create) can contribute to developmental unevenness.
4. Unexpected responses to writing tasks by student-writers have inevitably occurred no matter how carefully test makers have tried to control for

provided, and when. If a prompt is intended to be open, prompt authors should avoid obscured contexts which may lead some readers (such as many of the students in this study) to attempt a reconstruction of the authors' intended frame of reference. If a context of prior knowledge is considered to be important, as, of course, it is in many assessment situations, then prompts should clearly specify what this context is."

8. "When students have only one chance to write and are being judged on this single performance, it is especially important that they have optimal conditions for their performance, and that there be no mismatches of either topic or task."
9. "Various interpretations were made of prompt segments, and these interpretations were related to rhetorical strategies adopted by the students in composing a response."
10. "The first step for both the student and the evaluator in a writing assessment is the reading and interpreting of a prompt. Thus, the features of a prompt which influence interpretation are of critical importance for both the reliability and the validity of our assessment measures."

#### C. Issues and Recommendations

1. Further study is needed of "possible within-topic salience," or the possibility that certain individual words in the prompt might cue certain individual students more strongly (because of the effect of the student's prior knowledge on his or her response to certain words).
2. Disagreement emerged between students and teachers as to whether or not the writing produced by the prompt would give a good indication of writing ability.
3. "The study's findings have serious implications for current practice in assessment, since mismatches of the kinds described can result in distortions of the ways individual students are judged."

such responses.

5. Instead of continuing to narrow task constraints in the writing assignment as a way to control for unexpected outcomes, evaluators should attempt to describe what students do differently from one writing occasion to the next, interpret that difference, and attempt to measure the development it implies. (This is what was attempted on the Dřaké longitudinal study, using the typology described here.)

### C. Issues and Recommendations

1. "Empirical research is needed to help discover why a student construes a task as she does: what is the range of options she has? How does she want her text to function and why did she set that goal? What strategies will she choose to reach her goal? Is she an oldtimer or a newcomer to her chosen strategy? What awarenesses of text grammar appear to influence her whole text planning or her sentence level planning? How is the onset of text level planning related to increasing awareness of discourse schemas? The discourse typology, refined and clarified, may be of some use in exploring these questions, questions which may help us discover why and how students change the difficulty of what they attempt, even when a tester may intend task demands to be parallel. These questions may also help us define task complexity in new ways, allowing more precise evaluation of changes or growth in composing ability."
2. "Much greater clarification of discourse features, and of the role played by text level expectations of both writers and readers, seems needed before it will be possible to measure improvement in composing abilities".

## Part II: The Model in Practice

Elissa Warantz and Catharine Keech

### A. The Problem

1. A longitudinal examination of "the extent to which students attempt to do different kinds of writing, either spontaneously or in response to strongly suggestive features of the writing assignment, in a series of annual holistic writing tests taken during their ninth to twelfth years in high school." The hypothesis was that the differences in what the student attempted--the task construction--might help account for "unexpected directions of change in student performance scores from year to year." Results of this research would contribute toward an understanding of "the seeming instability of student performance from one test occasion to another," directing attention to the "developmental features of text design and task analysis which may influence the quality rating of student essays in seemingly paradoxical ways; i.e., when student test scores seem to deteriorate over time, rather than improve."
2. Although test makers can control the set of constraints and options expressed or implied in the text of the writing assignment (the given task) and thus create parallel task demands on subsequent occasions, they are less able to control the student's construal of the task, which determines the set of constraints actually honored and the options actually chosen by the student in writing (the constructed task).
3. Prior to any effort at measuring the development of writing ability, what is needed is "a method of comparing what students actually do and how they do it, from one occasion to the next." Therefore, the first step in this research was "the derivation of a typology of task construction which would adequately account for the range and variation of writing produced by students, as well as those texts traditionally classified in terms of literary 'genres.'"

## B. Outcomes

1. The typology derived was applied to a sample of student papers and the resulting data provided a way to "compare individual student performance (in terms of task construction and score) to the favored task construction and mean score of all students in a given topic/year."
2. The most frequent source of discrepancy in coding of functions and strategies (a process that the teacher-raters were trained to carry out with these papers) appeared in papers which, according to the raters, seemed to contain some combination of description and evaluation in function and/or strategy.
3. Discrepancies, along with the general ambiguity in the final coding of such papers, seemed attributable to two factors: (1) topic effect, and (2) the "bleeding strategy effect" on rater perceptions.
4. Another relatively large group of papers for which the typology could not adequately account were coded as "unclassifiable" in terms of function or strategy or both. These papers "appeared to fulfill more than two functions and/or strategies simultaneously (and therefore contained no clearly defined primary function or strategy), or they fulfilled a social or phatic function...which overrode any rhetorical function included in the typology."

## C. Issues and Recommendations

1. "Only after we are able to describe what the student actually produces, in response to test prompts, will we be able to infer the student's conceptualization or construction of the writing task. Then we will be able to generate and test hypotheses about the relative level of difficulty of what is attempted. That research should, in turn, make possible a proper investigation of the development of writing skills, by allowing us to distinguish between a student's mastery of old tasks and his novice attempts to produce new tasks. Further, such research may ultimately allow researchers to predict which kinds of tasks are best learned first, which follow, and what the various spontaneous task constructions employed by students may indicate about a given student's particular level of development."
2. After determining which task constructions were favored for a given grade within a given topic/year, further analysis would reveal whether a

student's use of an "unfavored" task construction  
in a given year might have affected that student's  
score.

## REFERENCES

- Bereiter, C.; Scardamalia, M.; and Bracewell, R. J. (1979). Quoted in Polin, Linda. Specifying the writing domain for assessment: recommendations to the practitioner. (CSE Report No. 135) Los Angeles: Graduate School of Education, University of California, Los Angeles 1980.
- Coffman, William E. Essay examinations. In Educational Measurement. Washington: American Council on Education, 1971.
- Crowhurst, M.C. Syntactic complexity in two modes of discourse at grades 6, 10, and 12. Washington, D.C.: U.S. Department of HEW, 1978. (ERIC Document Reproduction Service No. ED 168 037).
- French, (1966). In Foley, Joseph. Evaluation of learning in writing. In Benjamin S. Bloom (Ed.) Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.
- Godshalk, F.I., Swineford, F. and Coffman, W.E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.
- Greenberg, Karen L. Some relationships between writing tasks and students' writing performance. Paper delivered at the 71st Annual Convention of the National Council of Teachers of English, Boston, 1981.
- Hoetker, James. Effects of essay topics on student writing: a review of the literature. Tallahassee, Florida: Department of Education, Florida State University, 1981.
- Kinneavy, James L. A theory of discourse. New York: W.W. Norton and Company, 1971.
- Lloyd-Jones, Richard. Primary trait scoring. In Cooper, C. and Odell, L. (Eds.), Evaluating writing: describing, measuring, judging. Urbana, Ill.: National Council of Teachers of English, 1977.

- Mellon, John. National assessment and the teaching of English. Urbana, Illinois: National Council of Teachers of English, 1975.
- Michigan Survey Research Center. Interview manual. Michigan, 1969
- Odell, Lee. Teachers of composition and needed research in discourse theory. College Composition and Communication, 1979, 30, 39-44.
- Odell, Lee, Cooper, Charles R. and Courts, Cynthia. Discourse theory; implications for reearch in composing. In Research on composing: points of departure. Cooper and Odell (Eds.), Urbana, Illinois: National Council of Teachers of English, 1978.
- Payne, David A. and McMorris, Robert F. Educational and psychological measurement. Waltham, Massachusetts: Blaisdell Publishing Co., 1967
- Polin, Linda. Specifying the writing domain for assessment: recommendations to the practitioner. (CSE Report No. 135) Los Angeles: Graduate School of Education, University of California, Los Angeles, 1980.
- Sax, Gilbert. Principles of educational measurement and evaluation. Belmont, California: Wadsworth Publishing Co., 1974.
- Smith, Laura Spooner. Measures of high shcool students' expository writing: direct and indirect strategies. (CSE Report No. 133), Los Angeles: UCLA Graduate School of Education, 1980.
- Winters, Lynn. The effects of differing response criteria on the assessment of writing competence. (CSE Report No. 131), Los Angeles: UCLA Graduate School of Education, 1980.

Chapter 1

DESIGNING PROMPTS FOR HOLISTIC WRITING ASSESSMENTS:  
KNOWLEDGE FROM THEORY, RESEARCH, AND PRACTICE

Leo Ruth and Catharine Keech

Part I: Sources of Knowledge for Designing Writing  
Test Prompts

Leo Ruth

Part II: Practices in Designing Writing Test Prompts:  
Analysis and Recommendations

Catharine Keech

## Chapter 1

### PART I: SOURCES OF KNOWLEDGE FOR DESIGNING WRITING TEST PROMPTS

#### OVERVIEW OF THE STUDY

Often the first question asked in schools embarking on an assessment of writing is, "What topic shall we use and how should we word it?" An enterprising Coordinator of Competency Assessment for a California high school offers the following advice in an article on "Organizing for Holistic Scoring" in California English (Paulin, 1980):

Develop a list of promising topics--as many as you can find, borrow, or lift from other assessments.  
Write up directions to students for each topic.

Is there no better advice to offer prospective evaluators than simply to pillage from previous tests? Is there some rational procedure for the generation of writing test prompts? Are there systematic and theoretically sound rules for designing appropriate writing tasks for local, state, or national assessments of writing?

These are important questions to seek answers for, because it is the writing test prompt that initiates a whole process culminating in a judgment of writing competence that presumes to say something meaningful about the writing experience of the student in relation to the instruction received. Thus, the conception of a writing task implies a model of discourse forms and functions and a theory of the nature of the writing process. Inasmuch as the prompt must stand in dynamic relation to the piece of writing it initiates, prospective evaluators presumably would desire to learn how the various properties of writing prompts might affect the performance of writers in both intended and unintended ways. So if our California Coordinator of Competency Assessment were to look to those authoritative sources that school people are so often adjured to consult, what useful knowledge might she uncover?

The Research Problem. The principal research purpose governing the work of all dimensions of this NIE/BAWP Writing Assessment Project is to increase our knowledge of the properties of writing test prompts and the compositions they initiate. There is ample evidence in daily experience to enable us to conclude that particular prompts do not perform as uniform stimuli across populations to guarantee the activation of expected levels of thought and forms of discourse. This investigation asks the questions:

1. What elements in the content and structure of writing test prompts influence the production of compositions during writing assessments?
2. What makes a prompt "good" for eliciting writing samples to be rated under conditions of "holistic scoring" procedures?
3. How are "good" writing prompts generated and tested systematically?
4. What changes occur in maturing students' responses to writing prompts in successive annual assessments, and what do these changes reveal about how we should be assessing developing abilities in writing?
5. What sources of knowledge do we have to guide current practice in designing effective writing prompts?

This chapter addresses questions 2, 3, and 5, emphasizing primarily a description of the sources of knowledge from theory, research and practice.

William Irscher, currently president of the National Council of Teachers of English, observes, "...it is hardly possible to discover new ground if one does not even know what is established territory." Therefore, the place to begin to seek answers to the first two questions above is with a review of the research. To date, there appears to have been only one other comprehensive review of research on writing topics (Hoetker, 1980), described more fully later. The review envisioned here will reach out for new ground not covered by Hoetker, especially in the realm of retrospective generalizations collected from professionals--teachers, scholars, researchers, professional evaluators. Taken together their collective practical wisdom coalesces into a kind of "theory of practice" to be discussed later. The more conventional sources--psychometrics, for example--also are consulted.

Following the extended review of the literature, a second phase of this study of "who knows what about designing good writing prompts" will describe a procedure whereby we circulated a draft of a report on practices in designing writing test prompts among two panels of reviewers: one was a cadre of teacher-consultants from the Bay Area Writing Project and other writing instructors; the other was a group of writing researchers and professional test-makers. The responses of the two panels are collated and reported later in this chapter, and the revised report on practices in designing writing test prompts constitutes Part II of this chapter.

Limitations of the Study. Because the Bay Area Writing Project has pioneered the dissemination, through its many National Writing Project sites, of a "holistic writing assessment" procedure, this study is limited to an examination of those kinds of prompts which may serve the purposes of this model. Thus, this study excludes the well-regarded "primary trait" system of the National Assessment of Educational Progress. The primary trait system requires uniform specification of a set of task constraints in order to "evaluate the capacity to write for precisely defined purposes," according to the most recent NAEP Procedural Handbook, 1978-79 Writing Assessment. A fuller discussion of this model is provided later in this chapter, but the type of writing "exercise" derived from the primary trait model is not treated in any great detail. Other scoring systems--analytical, feature analysis, essay scales--which may also entail certain considerations in the design of writing prompts are not considered here.

Neither is the principal instrument of ongoing writing assessment, the class writing assignment, considered here. The class writing assignment is one of the areas of investigation covered in another BAWP/NIE project: "Research on Effective Writing Instruction in Inner City Secondary Schools." Class assignments also frequently are the subject of project publications such as the National Writing Project series on Writing Teachers at Work, e.g., Ray Marik's Special Education Students Write: Classroom Activities and Assignments (University of California, Berkeley: NWP, 1982). There is also the series of Classroom Research Studies done by three classroom teachers--Patrick Woodworth, Rebekah Caplan, and Stephanie Gray in association with Catharine Keech as the research design consultant. These publications are listed and described in a Publications for Teachers list appended to this chapter. It is necessary, however, to take a moment to distinguish between the nature of class writing assignments and school writing tests.

Class Writing Assignments and School Writing Tests Distinguished. Because it has been frequently suggested that our focus on the study of writing assessment prompts be expanded to include classroom writing assignments as well, it is necessary to distinguish carefully between the character of these two types of occasions for writing. The prompt that appears in a testing context must function autonomously: it permits no aid in interpretation from the proctors of the event; it offers no specific preparatory context or motivating impulse other than the command to write. The conditions for the participants are carefully controlled in order to be depersonalized, formal, standardized. The underlying psychometric assumption upon which the stability of such measurement relies is the belief in the uniformity of the stimulus properties for each student responding to the prompt. Each head bent over the page is presumed to be getting the same message to direct his/her writing performance.

By contrast, the prompts (or writing assignments) that appear in a classroom instructional context function in a setting that encourages negotiation of interpretation. The teacher may offer a good deal of guidance, making clear the expectations as well as the possibilities of the assignment. The teacher may orient the class to the area of experience to be drawn upon and, through discussion, may enlarge the ground of shared knowledge. The teacher may give examples of appropriate and inappropriate strategies for discovering and structuring the essential material of the composition. And, if the writing occurs in class, the teacher can assist in repairing any of the breakdowns that may happen if initial attempts to progress through the writing assignment fail. The process is personal, more informal, unstandardized and keyed to the proposition that differential interpretations of the prompt are inevitable but also reconcilable through negotiation. A formulation of major distinctions between the use of writing prompts in testing occasions and in teaching occasions follows:

<u>PARTICIPANTS AND PROCESS</u>	<u>WRITING ASSIGNMENT IN A TEACHING CONTEXT</u>	<u>WRITING PROMPT IN A TESTING CONTEXT</u>
1. Population:	Class members	Assembled students
2. Leader:	Teacher	Unknown proctor
3. Assignment Text:	Negotiable	Autonomous
4. Interaction:	Allowable	Prohibited
5. Writing Process:	Interpersonal Reciprocal Informal	Depersonalized One-way Standardized
6. Expectations	Known	Partly implicit
7. Evaluation:	Flexible Contingent Graded	Fixed standard Ranked Scored

The two types of assignments must meet different criteria to perform their unique functions. Thus assignments for teaching and testing are not easily interchangeable. Keeping these sorts of distinctions in mind, we shall now examine the several sources of knowledge we have available to inform practice in designing writing test prompts. The chief sources to be consulted in the following section are psychometrics, professional wisdom, discourse theory, and research.

SOURCES OF KNOWLEDGE FROM THEORY AND RESEARCH  
FOR DESIGNING WRITING TEST PROMPTS

In the skillful framing of a writing task, more than in anything else, lies the fine art of assessing writing. In the writing task the student finds the spur to imagination, the stimulus to thought, the guide to idea structure, and the incentive to writing performance. The writing task set in an assessment literally "prompts" the writing performance evoked. Inasmuch as the statement of the task stands in dynamic relation to the piece of writing it initiates, examiners presumably would desire to know how particular properties of tasks might affect the performance of writers both in predicted, and in unintended ways. Given the long history of the use of written examinations in this country, one might expect to find a body of knowledge drawn from theory, research, and practice which guides the design of writing test questions, assignments, tasks, or prompts. After all, the written examination emerged in the middle of the nineteenth century in the United States and thereafter became the dominant means of assessing learning. In the mid-thirties, the very term examination was still synonymous with written examination. Thus, a report on the subject to the Carnegie Foundation for the Advancement of Teaching was called Examinations and Their Substitutes in the United States (Kandell, 1936). In this report the "substitutes" for "real" examinations were the "new type tests" which provided several varieties of "objective" questions.

As the claims of superiority for the "new type" test became stronger, efforts were being made to formulate essay questions in such a way that they too could be marked more objectively and yield more reliable results. But as Kandell pointed out, "...the more accurately and carefully this is done, the closer will be the approximation to the new type test and the less in the long run will be the real educational value of the essay" as a matter of training in expression (p. 85). In 1936, according to Kandell, the leading advocates of the "new type" test were not yet suggesting that it should replace the essay as a "form of training and practice in clear organization and logical presentation of facts, concepts, and ideas" (pp. 84-85).

Nevertheless, Kandell did foresee that

...there is a real danger that, because of the greater reliability and validity of sound objective tests, the educative value of writing essays and their value as a method of training in

expression, in assembling materials, and in organizing them logically may be neglected, or, if not neglected, be relegated to classes in English composition alone (p. 68).

Now that the "new type" objective tests have become the dominant means of assessing writing performance, it is difficult to imagine the time when written examinations were the chief means of taking a "daguerreotype likeness...of the state and condition of pupils' minds" (Horace Mann quoted in Kandell, p. 26).

Until the turn of the century, the discussion of the properties and effects of written examinations was still based on opinion (Kandell, 1936, p. 57). But as the "scientific movement" in education began to grow, pioneers in educational measurement such as J. Mckeen Cattell and Edward L. Thorndike began to apply statistical methods to the study of written examinations. In 1905, Cattell declared that "it seems strange that no scientific study of any consequence has been made to determine the validity of our methods, to standardize and improve [written examinations]." In this same paper, Cattell staked out the territory that was to become the contemporary field of psychometrics when he prophetically argued

It seems scarcely possible to determine whether students are fitted for a college course by means of a written examination.....To devise and apply the best methods of determining fitness is the business of the psychological expert, who will probably represent at the close of this century as important a profession as medicine, law or church (quoted in Kandel, p. 59).

Cattell's challenge laid the foundation for subsequent study of the reliability of essay examinations and the development of their "substitutes," objective tests.

#### Psychometrics as a Source of Knowledge

Psychometrics has dealt with a restricted range of concerns. As Gilbert Sax defines the field: "Educational measurement...requires the quantification of attributes according to specified rules" (1974, p. 3). These rules deal with the measurable properties of tests to resolve issues of reliability, validity, objectivity, and standardization in the administration, scoring, and interpretation of tests. While textbooks in measurement do deal with test construction in a general way, their focus tends to be more on the technology and methodology of measurement than upon the substantive properties of the items or the essay questions in a test. That is to

say, measurement textbooks provide methods of applying statistical approaches to the appraisal of the "behavior" of writing prompts or objective test items during field trials, but they do not offer much beyond the most general sorts of principles for actually designing and wording the writing tasks. The field trial provides a self-evaluative procedure for eliminating faulty essay questions or other test items, but it does not reveal what has gone wrong in framing the item. The procedure is simple: to design an essay examination that will distinguish between students who do well in writing and those who do not, the examiners simply compose several essay questions, administer them to a sample of the population to be tested, observe whether they predict performance along the lines expected, and eliminate or change questions that do not work. The essay questions that remain are presumed to be "meaningful" to the larger population to be tested. Psychometrics strives to be scientific, but the procedures used to construct and validate essay questions are essentially subjective and scientifically crude in the sense that there is no theory of information processing underlying the design of writing tasks. Standardized testing procedures presuppose that the examiner and the student writer share a common language and uniform interpretations of tasks. But the function of a writing task as a stimulus event cannot be understood until the examiner understands clearly how the constraints of language use and interpretation operate in the writing assessment event.

Statistically oriented evaluators do not seem to be aware that the psychometric approach to assessments of ability necessarily presupposes an information processing model of communication. Consequently, psychometricians have not attempted to study the influence of the psycholinguistic and sociolinguistic variables on the process of interpreting objective items and essay questions used in particular writing assessment contexts. Rather, the emphasis in psychometrics has been on the manipulation and treatment of products in measurement schemes, not on the understanding of the cognitive processes that might also be involved. From the beginning, psychometrics has been interested in concrete matters in the measurement of an educational product. As defined by Edward L. Thorndike in 1918, educational measurement is based upon the following theory:

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality. Education is concerned with changes in human beings; a change is a difference between two conditions; each of these conditions is known to us only by the products produced by it--things made, words

spoken, acts performed and the like. To measure any one of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. (Thorndike's 1918 statement is in the National Society for the Study of Education, Seventeenth Yearbook, pp. 16-17, as quoted by Kandel, p. 80).

Given this notion of measurement, it is not surprising that Thorndike and other early leaders in measurement devoted themselves to reducing the weaknesses and defects in written essay types of examinations by devising "new type" tests which yielded results that were objective and were considered more reliable and accurate.

I.M. Kandel, reporting to the Carnegie Foundation on the state of knowledge in conducting examinations, observed that investigations had been "limited in the main to a study of the reliability of the marking of examinations" (Kandel, p. 3). Thirty-five years later, in writing the chapter on essay testing for the definitive volume, Educational Testing (1971), published by the American Council on Education, William E. Coffman, a researcher for the Educational Testing Service and the College Entrance Examination Board, could still confirm Kandel's earlier observation about the dominance of issues of reliability in the study of essay testing. Coffman says that most research on essay testing "has dealt in one way or another with the question of reliability" (Coffman, 1971, p. 298). The most recent review of research on the effects of topics on student writing now available also supports the findings of the two previous reviewers. After conducting his comprehensive review, James Hoetker (1981, p. 4) stated that "research attention has been devoted almost entirely to issues of rater reliability, ignoring for the most part the issue of validity as well as the other two sources of error in an essay examination--the topics and the writer."

Psychometric literature does recognize the possibility of variation in the difficulty of writing tasks, but it views the issue broadly and non-analytically. There is seemingly little in the literature of measurement (see Sax, pp. 113-116) and educational research in the way of empirically determined principles for the formulation and wording of essay test questions. Such principles can, however, be found in the literature on instrument construction in public opinion research on the wording of questions for interviews and questionnaires. (See Payne, 1951; Oppenheim, 1966; Gordon, 1980). For example, Payne and McMorris (1967, p. 145), in an editorial note before one of the selections in their collection of readings on

the theory and practice of educational measurement, have commented as follows: "while textbooks dealing with instrument construction present many suggestions for item writing, the empirical evidence demonstrating the worth of such apparently sound rules is meager." The few available "rules" appear to be concerned chiefly with such aspects of item construction as equal length answer alternatives vs. extra long answer alternatives; consistency vs. inconsistency of grammar between lead and alternatives; the question type of stem vs. the incomplete statement type of stem, etc. (pp. 141-152). These "rules" obviously deal with matters of form rather than the wording of content. With regard to the wording of essay questions, Coffman advises the test-maker to make sure that essay questions are "so worded that all candidates will interpret in the same way the task to be done" (1971, p. 287). But Coffman offers no suggestions for accomplishing this goal other than to put the essay questions before colleagues for review or put them on trial in the field in a manner described earlier in this report.

"Field tests" and "item analysis" are the principal procedures now in use to determine whether an essay test question or an objective test item is well-constructed. Coffman recommends using field test results to determine whether:

- 1) the examinees understand the intent of the question, or whether they interpret it in unintended ways;
- 2) the question is of appropriate difficulty for the examinees; and
- 3) the question can be reliably scored.

In Coffman's view, the only way to get conclusive answers to these questions is to "obtain a representative sample of responses and conduct an experimental scoring session leading to the generation of score distributions and estimates of reliability" (p. 288). Coffman argues that even informal testing of an essay question with a small, unrepresentative sample of students is likely to be better than no pretesting at all, for "the contrast between the type of response anticipated by the examiner and the type actually written by examinees...can be dramatic" (pp. 288-89).

Interestingly, "item analysis" as it is practiced in psychometrics does not take into account the linguistic properties which might account for the ambiguity and difficulty of test questions. Rather the psychometric concept of "item analysis" refers to an application of statistical techniques to determine "difficulty" and

"ambiguity." The "difficulty" of a test item is "indicated by the proportion of the subjects who get the item right" (Wood, 1960, p. 81; see also, Sax, 1974, pp. 229-254). Item "ambiguity" from the point of view of psychometrics is also defined quantitatively according to the "extent which students in the upper group (top quartile) select an incorrect option with about the same frequency that they select a correct one" (p. 233). In other words, "ambiguity, is defined as the inability of the top scoring students in a test to discriminate between the "correct" and the "incorrect" alternatives as determined by the examiner. Sax (1974) acknowledges that the procedures of "item analysis" are such that "unfortunately [they] cannot reveal if ambiguity results from the students' lack of knowledge or from poorly written items" (p. 233). It is important to note that the statistical procedures used in the conventional field test of essay questions enable examiners to determine the task's difficulty and its "power to discriminate" between "successful" and "unsuccessful" students, but these procedures do not lead to an analysis of the linguistic properties inside the texts of essay questions which might affect their interpretation by the persons responding to them.

The interaction between students and essay test questions is recognized and demonstrated by psychometric research (Coffman, 1971, p. 289). For example, one investigator (French, 1966, p. 588, as cited in Foley, 1971, p. 801) identifies "test error" among the sources of error in scores based on tests requiring actual writing: "The composition test is almost like a one-item test. Some students may happen to enjoy the topic, while others may find it difficult and unstimulating; this results in error." Coffman (1971) goes so far as to say that "some students do better on some questions while other students do better on others. To some extent the grade a student obtains depends on which questions appear on examinations" (p. 289; see also Rosen, 1969). Despite the fact that researchers in educational measurement have acknowledged question-wording effects, the larger theoretical issues of essay task construction and typology have seldom been addressed. There has been little concern about determining the frequency, magnitude or underlying nature of essay task-wording effects. Wording effects have tended to be treated anecdotally or incidentally in the context of other objectives of study. For example, the fundamental research that is most often cited to confirm the validity of the use of multiple-choice questions as predictors of students' writing ability is a study done by Godshalk, Swineford, and Coffman

(1966).\* This study revealed significant variation in ratings assigned to the essays produced in response to the five topics used in the study, as reflected in the mean score for topics. The investigators point out that these differences in ratings would have consequences under certain examination conditions:

...if the five topics had been assigned as alternate topics from which one or two could be chosen by students, a student's rating might depend more on which topic he chose than on how well he wrote. Or if one topic had been assigned to one form of a test and another topic to a second form, then some method of equating the scores would be required; otherwise the magnitude of an individual's score would depend partly on which form of the test he wrote.

Though the topic effect is recognized as significant in the Godshalk study, it is considered only in connection with the estimation of the score reliability. Topic effect is identified as a source of variability in the reliability of the test score, but it is not discussed as an interesting phenomenon worthy of further investigation.

#### Professional Wisdom as a Source of Knowledge

Practitioner knowledge develops through a continuing involvement in the unique and particular experiences of the profession. The core of professional knowledge derived from the accumulation of insight through systematic self-inquiry leads to the fashioning of various degrees of practical wisdom. This practical wisdom may denote a high level of competence in practice, but it remains essentially an individual theory of practice so long as it stays private. However, there are individuals who make public their personally accumulated wisdom through a textbook on pedagogy, a curriculum report, or other media. Once public, these works are subject to critical processes which may lead to their verification, validation, and refinement. And, in time, we have the development of a form of theory that might be called "theory as idealized practice." This kind of theory contrasts with that other type of theory which is derived from systematic inquiry of the sort associated with formal research procedures.

It would require a separate research project to survey and systematically collect and synthesize available

\* Martin Steinman has challenged the conceptual validity of this famous study (Research in Teaching English, I (1), Spring 1967, pp. 79-94).

theories of idealized practice covering the design of writing tasks. Several examples from different periods in the history of teaching writing in the United States will suffice to indicate the type of normative "rules" recommended on the basis of practice. Each of the following sets of instructions pertains to making assignments for classroom instruction; however, they seem adaptable to an assessment context as well.

Fred N. Scott. In his "Philosophy of the Assignment," Fred N. Scott (1903)\* divides the making of an assignment into five steps:

1. the announcement of the subject;
2. stimulation of interest in the subject;
3. arousal of a desire to write upon it;
4. suggestion of a method or procedure in writing;
5. precautions against wasted effort.

Much of Scott's discussion is devoted to consideration of the principles for selecting an interesting subject. He considers the feasibility of allowing the student to select his own subject and grants that this is a valuable mental exercise. But in practice, freedom of choice does not work because too much of the effort expended is "mis-directed and desultory" (p. 321). He does, however, recommend carefully prepared special exercises in subject-choosing. Interestingly, Scott's professional judgment regarding subject choice anticipates psychometric research data which indicates that when students are given a choice of topics, they are unable to select the one upon which they will make the highest score (Coffman, 1971, p. 291).

Scott offers two main principles for choosing subjects for writing: 1) the subject chosen must be one that is interesting to the teacher, and 2) the subject chosen must be one that is interesting, or that can be made interesting, to the students (p. 322). The first principle is important because it reduces the possibility of topic effect upon the evaluator when the principle is observed. As Scott explains,

It must be borne in mind that after the essays are written they must be read, and it is the teacher who will read them. He ought to read with intense interest and a kindling enthusiasm. But if the subjects are distasteful to him, his reading, despite the most conscientious efforts, will be half-hearted and ineffectual. (p. 322).

In discussing the means of arousing the desire to write, Scott claims that "such a desire, when it is natural, springs from two healthy impulses: the impulse to give expression to one's thoughts and feelings, and the impulse to communicate one's thoughts" (p. 324). Again, Scott's professional wisdom anticipates

---

\* I am indebted to Alred H. Grommon, Professor Emeritus, Stanford University, for calling my attention to this source.

the arguments of latter-day discourse theorists such as James Moffett and James Britton, who also advocate beginning with assignments that elicit personal, expressive writing for real audiences. Scott stresses the importance of taking account of audience in making writing assignments: "...in assigning the work, the teacher should be at some pains to provide an audience or a reader. Sometimes the audience will be real, sometimes it will be imaginary, but it should never be lacking" (p. 325). Audience is an important consideration because, as Scott puts it, "the consciousness of waiting auditors" is "the most powerful of all stimuli to expression" (p. 325).

In summary, Scott states the goal of a well-designed writing assignment": "The effect of the whole should be to create a natural situation, real or imaginary, in which the student's powers of expression and communication are stimulated to their normal maximum" (p. 326).

Commission on English. In 1965, the Commission on English, an agency of the College Entrance Examination Board, published its report on the teaching of English, Freedom and Discipline in English. The Commission's report contains a section on "Composition Assignments" which declares, "No part of an English teacher's job is more important...than the making of sound, well-framed assignments, what is called 'providing the occasion'" (p. 92). This report established the following criteria for a good assignment (pp. 93-96):

A good assignment

1. evokes the best from the writer and gives the teacher the best chance to be helpful;
2. aids learning and requires a response that is the product of discovery;
3. furnishes data to start from;
4. may take the form of, or be construable into, a proposition;
5. limits either form or content or both;
6. will stipulate the audience to be addressed, wherever feasible.

The Commission proposes starting an assignment with data such as contradictory criticism of a literary text, opposing arguments, incongruous bits of common sense (e.g. contradictory maxims). The assignment may start from a picture or a cartoon, but often it should begin with the literature studied. The Commission recommends converting a topic into a proposition:

"The view from my window" has the virtue of inviting invention but the serious fault of giving the writer nothing to control his attention. Translated to propositional form--"That the view from my window makes

me dread (or welcome) getting up in the morning"-- the same topic suggests a focus and even a tone for the writer to exploit...The provision of a predicate immediately puts the writer into a posture of defense or attack and calls for the summoning and ordering of evidence or arguments. Instead of a circle circumscribing undefined matter, the proposition supplies an arrow pointing out a clear direction of movement (p. 94).

With respect to consideration of an audience, the Commission finds that "Too many English themes seem addressed...to the teacher...or, in a vague romantic way, to the world or posterity or some Saroyanesque 'You Out There'" (p. 95). The Commission's remedy for this lack of attention to audience is to have teachers vary the stipulations, beginning with fairly simple exercises: "Write a letter to the governor arguing that..., 'Write a petition to the student council or the principal requesting that ..., 'Write an essay on the proposition that..., 'etc." (p. 95). The Commission expects the student eventually to define his own audience and "to define it, not in so many words, but by tone and content alone" (p. 96).

The Commission does acknowledge that "for the novice" especially, it may be inhibiting rather than helpful to require that he think of a specific audience for a poem or for any of the range of forms that are more 'literary' than expository or argumentative" (p. 95). It can be seen at once that the Commission's writing assignment models actually go beyond naming an audience to specify also mode or genre and propositional content. Thus, these assignments circumscribe and define the essay that is to be written.

Josephine Miles. Josephine Miles believes that the solution to dealing with subjects of any magnitude lies with training the writer to formulate a "responsible predication" for any given topic. In her view, no subject is too large or too small or too complicated so long as the writer controls what he has to say about it. This control comes through the selection of a supportable predication about the subject from the writer's own experience or knowledge. Miles suggests that one can take even an abstract subject such as "death" and formulate a predication that is "responsible" and "supportable" in terms of what a writer knows: for example, "When death appeared in my life, it didn't frighten me" (p. 3). Miles also points to the structural implications of a good leading sentence. For example, the development of the main idea of the "death" predication calls for a retrospective account of an event as it happened. The temporal locative when in association with the sequence of verbs, appeared and then didn't frighten, together require a commitment to develop the natural temporal sequence they suggest. Miles believes that it is necessary for students to learn how to recognize the structure that is implicit in their own predications and in the questions they receive in assignments. Miles has formulated her theory on the basis of years of experience in teaching freshman English and through informal "experiments"

with teaching the cues to the structure of predications in statements and questions. She says:

Over the years...we have come to agree that assignment by topic is artificial, not only because it is unmotivated, but because it is too far removed from the central unit of thought, the statement. It is the question to be answered, rather than the topic to be predicated, which gives the student the central clue to the order and structure. Only after he can handle questions easily can he move easily to those mere half-questions which are topics (p. 20).

Josephine Miles criticizes the practice of offering assignments for writing which virtually exclude rational processes. What she thinks are most needed are opportunities for development of essays in ideas where writers can make "statements based on interest and speculation and [support] them by adequate evidence pro and con" (p. 23). The power of students to compose their thoughts comes through essays in ideas responsibly developed.

Commission on Writing. In 1966, the Commission on Writing of The Education Council proposed one maxim and four criteria for preparing composition questions:

Be sure that any question you ask insures the possibility of a rhetorically effective composition.

1. A good question should be stimulating: it should present a subject matter with which a student can readily become engaged and it should present the subject in such a way that he can make an assertion about it. Not "My Summer Vacation" but "What do you think was the point of failure in your summer vacation?"
2. A good question should be fashioned to the interest and abilities of the students. Like any writer, the poser of a question must be aware of his audience.
3. A good question should seek to elicit a specific response. Although one may at times wish a student to generate his own topic, generally one should seek to show the student the specific area in which he is to organize his essay and the audience for whom the essay is intended. A teacher should beware of his own power and not force the student to "write to the teacher."
4. A good question should be clear and precise in its instructions. In this respect, and in all respects, the demands on the questioner are identical with those on the writer (The Education Council,

William F. Irscher. In a manual for teaching expository writing, William F. Irscher (1979) draws upon his long teaching experience to say the things that he believes teachers engaged in the teaching of writing should consider. He offers a chapter on "Topics" which he opens with the sentence: "Assigning topics for writing is one of the most important things a composition teacher does" (p. 69). Irscher's discussion of topics echoes in many respects the principles espoused by the other professionals so far presented here. Like Scott, Irscher would not give certain topics to freshman students because he would not want to read the cliched essays that they would produce. In his view such topics "are not springboards for writing; they are traps. They don't give students a chance to reveal that they can think inventively" (p. 69). Irscher offers five criteria that a good topic ought to meet (pp. 68-72):

1. A good topic ought to have a purpose.
2. A good topic ought to be meaningful within the student's experience.
3. A good topic ought to prefer specific and immediate situations to abstract and theoretical ones.
4. A good topic posing a hypothetical situation should be within a student's grasp.
5. A good topic ought to encourage a student to write.

Irscher carefully explains that a "meaningful" topic is not necessarily a "personal" topic, though it should be involving. He observes that students do distinguish what they are willing to talk about in the dormitory from what they are willing to talk about in the classroom. That is, students consider some topics "too personal" (p. 70). Irscher explains also what he means by specificity in a topic by contrasting the abstract topic, "Discuss Freedom" with another one that is more immediate: "List the freedoms you enjoy where you live and the freedoms you are denied. What is the reason for the denials? Do you accept the reasons? Write an essay on the subject" (p. 70). And finally, Irscher considers the effects of topics which require writers to project themselves into hypothetical situations. He sees a topic like the following as an imaginative stimulus which implicitly invites satire: "If you were the first chimpanzee to be landed on Mars, what message would you deliver to Earthlings?" (p. 71). Hypothetical topics such as the following one provide an indirect vehicle for expressing thoughts that otherwise would be difficult to share: "Write an extended obituary notice for the newspaper upon the occasion of your own death" (p. 71). This kind of hypothetical topic actually concerns

the values that a student holds now even though it seems to project him or her into the future. Thus, Irscher finds it very "different from one that addresses a seventeen-year-old by saying, 'If you were an old man or woman...'" This latter type of topic is ill-advised, according to Irscher: "If the purpose is to find out what it is like to be old or how the old think, it would be far preferable to have young people talk to old people and report their impressions" (p. 71). Irscher also believes that choice is good but that "limitless choice ('Write on anything you want to') usually proves self-defeating. Students flounder when they can choose anything in the world (p. 72).

Edmund Farrell. One of the few articles to appear in the English Journal (pp. 428-431) exclusively devoted to a discussion of the making of composition assignments is Edmund Farrell's essay. Drawing upon his experience as a supervisor of student teachers in secondary English at the University of California, Berkeley, Farrell culled a number of writing assignments that he felt exemplified various sorts of lapses in making assignments. He found that:

Common to these assignments is the absence of a stipulated audience and/or purpose which would help the student to define himself in context, which would lead him to adopt an appropriate persona or 'speaking voice' in his composition (p. 430).

Farrell joins a long line of professionals who advocate specification of audience, role, and purpose in the writing assignment. But Farrell goes further than most other writers in suggesting patterns for creating assignments with built-in "selves" which function as cues to suggest the particular roles (and "voices") that the writer may adopt in relation to audience and purpose. Farrell offers four possibilities:

1. The "self" of an assignment may be internal to the selection, the purpose unspecified, and the audience private. (Example: Assume that you are Lady Macbeth and that you keep a diary. Write the five entries which precede your suicide.) (p. 430)
2. The "self" and the audience may be internal to the selection and the purpose specified. (Example: You are Tom and have been away from home now for three months. Write to Laura trying to explain to her why you left.) (p. 430)
3. The "self," audience, and purpose may be external to the selection. (Example: A friend of yours comes to you with a copy of Macbeth and says, "I understand you've read this play in class. What should I look for so I can most fully understand it?" What advice would you offer?) (p. 431)

4. The "self" may be internal to the selection and the audience and the purpose external. (Example: You are one of the inhabitants of Spoon River who has died. You have an opportunity to speak out from the grave, summarizing your life in a paragraph or two. What comments have you to make?) (p. 431)

Farrell believes that if students are given good assignments that enable them to adopt appropriate voices, "their writing should be more pleasurable to read and much easier to evaluate" (p. 431).

Arthur J. Carr. A professor of English at the University of Michigan, Arthur J. Carr made a kinescope of "Student Writing Assignment Based on 'Fire Walking in Ceylon,'" for the Commission on English in the 1960's. Carr also believes that the assignment significantly affects the quality of writing the students may produce in response to it. His kinescope production relates his experience with a particular writing assignment which had produced a terrible set of papers. But instead of berating his students for their obtuseness, he takes himself to task, asking, "What went wrong with my assignment?" After analyzing the things that have gone wrong in making the assignment (lack of preparation of the students, lack of clarity in the statement of the assignment, lack of a defined purpose for making the assignment), Carr delineates two basic assignment patterns. Broadly speaking, Carr sees these two kinds of assignments arising from a reading of the essay on fire walkers: one type would emphasize structure and form; the other, ideas. "If we restrict the choice of forms, then we should liberate ideas. If we restrict ideas, then we should liberate form." Carr believes in carefully selecting and controlling the number of constraints a student must cope with in a single assignment. He sees the making of assignments as something to which the true professional gives considerable thought. He contrasts the amateur and the professional:

...the amateur depends on inspiration, and this means that he may be victimized by his moods. As we become professional, we learn to carry out even the difficult parts of our profession expertly and regularly. Unlike the amateur, the professional applies principle and the results of objectively examined experience (p. 16, 1965).

A Professional Consensus on Principles. So far we have considered the practical wisdom of several professionals situated at various levels of instruction with regard to their recommendations for designing writing assignments for classroom instruction. It is not known whether this group can stand as representative, for no empirically derived sample has been drawn from the universe of professionals with published works on creating classroom writing exercises. Nevertheless, to add more cases here would overwhelm. The thirteen members of the Commission on English and seven members of The Education Council plus the five other persons presented in this section together do constitute a noteworthy professional

community (see Appendix for institutional affiliations). Their collective wisdom seems to reach consensus on several principles for designing classroom assignments that can be transferred to designing prompts for writing assessments. Consider the following set of principles.

#### THE PHILOSOPHY OF THE ASSIGNMENT FOR WRITING ASSESSMENT

**Maxim:** A well-framed assignment for a writing assessment provides an occasion for writing in which the student's powers of expression and communication are stimulated to their normal maximum.

**Principles:**

1. The subject chosen must be potentially interesting to the teacher-reader or evaluator of the essays written.
2. The subject chosen must be potentially interesting to the student writers.
3. The assignment furnishes data to start from which is open to presentation in any of several forms ranging from statements or questions to pictures or cartoons accompanied by instructions.
4. The assignment must be meaningful within the student's experience.
5. The assignment seeks to elicit a specific response and limits content or form or both.
6. The assignment is itself specific in suggesting an audience beyond the teacher or evaluator when the subject makes such specification feasible.
7. Assignment by topic alone is artificial and yields a lack of focus in development.

**Coda:** A well-framed assignment, clear and precise in its instructions, shows awareness of the audience of waiting examinees. The demands on the assignment creator are identical with those on the writer.

Marjorie Kirrie. It is interesting to contrast this composite set of principles for making writing assessment prompts to some recommendations offered by a professional who is a teacher and a chief reader for the College Entrance Examination Board. In a brief article for the National Writing Project Network Newsletter, Marjorie Kirrie (1979) explains why it is necessary to distinguish between the instructional and the assessment situation in conceiving of satisfactory prompts. Since this distinction has

already been discussed before, we will proceed to her main points.

In a well-intentioned effort to be specific, prompt writers are likely to unwittingly introduce factors that produce unwanted side effects. These side effects are the result of what Kirrie calls "overwriting" the prompt. She cautions against "shoring up topics with all kinds of directions: 'You are writing to..., 'You have met a..., 'Tell about..., 'Tell how...'" (p. 7). Such instructions dictate structure and lead to "a stringing together of attempts to cope with each 'Tell' in the exact order given in the prompt" (p. 7). Kirrie also cautions against asking students to address a specific unreal audience, to imagine themselves in situations alien to their world, or to imagine themselves as other entities. The prompt that begins "If you were..." can only be written in the subjunctive or conditional modes and sets "traps for all except the highly skilled" (p. 7). Unlike the other professionals quoted earlier, Kirrie maintains that specifying an audience makes the writing task needlessly difficult and calls for greater skill than many students possess. "Besides," says Kirrie, "students never forget the real audience for assessment writing [teacher and other educators]" (p. 7). Also, Kirrie notes "some audiences call for less sophisticated diction and syntax than we would like to see students produce in a test situation" (p. 7).

So what does a prompt author need to do in order to make a prompt that works? Kirrie advises keeping the prompt as "non-directive as possible" and using "any word, phrase, or brief statement which invites a variety of interpretations and responses" (p. 7). She says:

A statement should give students as many options as did the quotation from Pogo used in last year's College Entrance Examination Board Test essay in (1978), "We have met the enemy and he is us," which elicited everything from historical exposition to commentaries on language usage. As for additional directions, variants of "Write an essay on what...means (suggests) to you," together with the advice that examples may be helpful and that good writing will be appreciated, are all that are needed (p. 7).

NAEP's Primary Trait System. While Marjorie Kirrie represents the professional view that holds that a writing prompt should have "the widest possible evocative range" (p. 7), there is another view that is diametrically opposite to this one. The National Assessment of Educational Progress's "primary trait" approach calls for a precisely defined writing "exercise." Ina Mullis (1976) of NAEP has described the characteristics of writing exercises devised for assessments using a primary trait system:

The essence of the primary trait system is to narrowly delineate the situation of the writer, by defining the variables. With this approach responses should all address the same task and can be judged using the same

criteria. Systematic judgments provide both descriptive and comparative information about how well individuals or groups of people can use writing to communicate in given situations. This means that for each exercise three things must be specified: (1) the identity of the writer (whether the respondent is himself or is given a role to play), (2) the audience (who the writer is writing to) and (3) the subject matter (what the writer should communicate to the audience).

The more structured the task, the less difficult the scoring, since the essays or letters will be more uniform in focus. For example, if persuasive writing is chosen as an important skill to measure, it could be decided that students should be able to use writing to influence decision-makers. The task could be, "Write a letter to someone important about a problem in this country." A better task would be, "Write a letter to your principal suggesting a way to solve a problem in your school." If the nature of the problem is further defined, e.g., the lunchroom, the task would be even better. With a national sample, it is difficult to identify universally applicable situations. For a classroom, school, or even a district, the problem of respondents having common knowledge and experience should be alleviated (pp. 8-9).

Rex Brown (1978), also a member of the NAEP staff, presents some ideas about how to develop an ideal writing assessment instrument. He also speaks on behalf of the structured writing task:

If you want to evaluate an essay for certain characteristics, then you must be sure that you have requested them in the assignment. This is not a trivial matter: it is extremely difficult to write assignments that define precisely the rhetorical imperatives that will either be met or missed by the students. If you want to know whether they can elaborate upon a role expressively while maintaining control of a point of view and tense then you have to set the task up in such a way that they must do so, and define acceptable levels of achievement that are concrete and realistic (p. 5).

#### Discourse Theory as a Source of Knowledge

Discourse theory, which deals with the development of full texts in speech or writing (Kinneavy's definition, 1971), provides another source of knowledge that influences the design of writing assignments for both instructional and testing occasions. Discourse theory with roots and traditions reaching back to antiquity attempts to describe the varieties of discourse and the elements which contribute to the making of texts, spoken or written.

Various schools of thought have emphasized one or another element such as the writer, the subject, or the audience, either individually or taken together. A fourth element, function (also called purpose or aim) arises from the interaction of writer, subject, and audience. Related to all of these is the concept of form (also called mode or genre). The basic problem in dealing with this field is the very richness of the area and the problems attendant upon selecting from among competing theories. Kinneavy (1980) sees the current condition in discourse theory and rhetoric as comparable to the vigorously competitive scene at the end of the 1950's when various schools of linguistics were vying for dominance (p. 37). It is outside the scope of this review to treat in any depth even a very selective sample of prominent theories. But it is important to consider in brief outline several theories which often provide organizing principles for instruction in writing. These various discourse schemes categorize forms and functions of writing that frequently turn up as specifications and constraints in writing prompts.

Discourse theory has gone in two directions. Traditionally, discourse theorists have looked at the products of writing and sought to classify what they found. More recently, the theorists have attempted to describe the processes of writing as well as the works created. The scope of discourse theory, including significant traditional and emerging contemporary models, is presented in Kinneavy (1971) and D'Angelo (1976) although neither of these authors considers the work of James Britton, Janet Emig, and Linda Flower and John Hayes, who have all proposed process models of discourse.

Alexander Bain. D'Angelo (1976, p. 115) credits Alexander Bain with having established in 1890 the traditional forms of discourse which have dominated the organization of writing curricula and testing even to this day. The traditional forms are description, narration, exposition, and argumentation, although Bain's original set also included a fifth form, poetry. According to D'Angelo's account of Bain's theory:

Each form is assumed to have its own function, its own subject matter, its own organizational patterns, and its own language (p. 115)....

...the mind can be divided into three faculties: the understanding, the will, and the feelings. The aims of discourse (to inform, to persuade, and to please) correspond to these three faculties. The forms of discourse are the kinds of composition that relate to the faculties of the mind, the aims of discourse, and the laws of thought. Thus, description, narration, and exposition relate to the faculty of understanding, persuasion relates to the will, and poetry to the feelings (p. 116).

Bain's system classifies and characterizes written products. Graphically portrayed, his distinctions are as follows (based on D'Angelo, p. 115):

	<u>Function</u>	<u>Subject</u>	<u>Organization</u>	<u>Language</u>
Description	Evoke sense experience	Objects of senses	Space/time	Denotative and connotative, figurative, literal, impressionistic, objective
Narration	Tell a story, narrate an event	People and events	Space/time	As above
Exposition	Inform, instruct, present ideas	Ideas, generalizations	Logical analysis and classification	Denotative and factual
Argument	Convince, persuade, defend, refute	Issues	Deduction and induction	Factual and emotive, depending on appeal

Leo Rockas. A Contemporary approach to the classification of written products is offered by Leo Rockas (1964) as he seeks to combine the literary modes identified by Plato and Aristotle with the rhetorical modes identified by Bain. Rockas says:

My analysis has suggested that these poetical and rhetorical modes extend themselves naturally and elegantly to the concrete modes, description, narration, drama, and reverie; and to the abstract or plagal modes, definition, process, dialogue, and persuasion. These are also classified by means of procedure as the static modes, description and definition; the temporal modes, narration and process; the mimetic modes, drama and dialogue; and the mental modes, reverie and persuasion. Though the modes may be modulated and mixed, I consider them to be exclusive of each other, and together inclusive of whatever can happen in discourse, at least at the simplest level of rhetorical analysis (p. ix.).

Displayed in graphic form, Rockas' reclassification appears below (pp. xii-xv):

- I. The static modes
  - A. Descriptions
  - B. Definition
- II. The temporal modes
  - A. Narration
  - B. Process

III. The mimetic modes

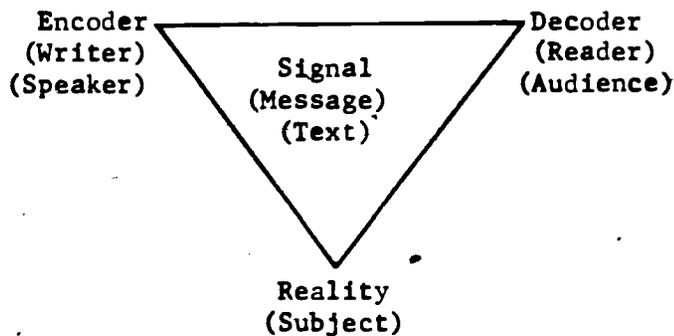
- A. Drama
- B. Dialogue

IV. The mental modes

- A. Reverie
- B. Persuasion

The A categories represent the concrete modes while the B categories constitute the abstract modes.

James Kinneavy. Still another scheme for classifying the modes and aims of discourse was proposed by Kinneavy (1971) in a comprehensive, multi-disciplinary examination of the contributions of classical and contemporary rhetoric. Kinneavy begins with the communication triangle formulated in communication theory and divides the universe of language into three fields of study (syntactics, semantics, and pragmatics) which are related to the terms of the communication triangle (pp. 19-31). The communication triangle is depicted below (with annotations in parentheses):



The area most closely tied to discourse in Kinneavy's scheme is pragmatics, which concerns the study of the use of language in actual speech (or written) situations (p. 22). According to Kinneavy the aims of discourse are based on the four elements of the triangle. When the focus of communication is primarily on the speaker-writer and the aim or purpose is self-expression, the resulting text is expressive discourse. When the focus is on the reader/audience and the speaker/writer's aim or purpose is to persuade, the resulting text is persuasive discourse. When the focus is upon the subject or reality under consideration and the speaker-writer's aim is to explain the world clearly and logically, the result is referential discourse. When the focus is upon the text and the speaker-writer's aim or purpose is to give pleasure, the result is literary discourse. To summarize: "Language can therefore be employed with the stress of the process on the persons (encoder or decoder), or the reality to which reference is made, or on the product (text which the discourse produces)" (Kinneavy, p. 38).

Kinneavy sees the modes of discourse as "grounded in certain philosophic concepts of the nature of reality considered as being

or becoming" (p. 36). Thus, the modes refer to classifications of "kinds" of realities. But Kinneavy alters Bain's traditional quartet, keeping description and narration but substituting evaluation and classification for argument and exposition. Kinneavy rejects exposition and argument as modes because he considers them aims of discourse (1980, p. 49). In other words, Kinneavy uses the term "mode" to cover "what" is being talked about, and "aim" to cover the "why" or the purpose or function of what is being talked about.

Kinneavy (1980) summarizes his own theory of discourse and compares it to the theories of James Moffett, James Britton, and Frank D'Angelo. In reaching for a "pluralistic synthesis" of four contemporary models, Kinneavy identifies a number of compatibilities between his own model and those of the other three writers. Kinneavy points out that three of the models (Moffett, Britton, and Kinneavy) are "explicitly based on the semiotic structure of the relationships among writer, reader, and subject matter" (p. 38). He then goes on to note a great many other points of congruence among the four models, but he does acknowledge that "the developmental dimension found both in Moffett and Britton is entirely lacking in Kinneavy and D'Angelo" (p. 47). Because they do address the developmental dimension of composing, both Moffett's and Britton's theories will be presented briefly below.

James Moffett. The initial presentation of James Moffett's discourse model appeared in Teaching the Universe of Discourse (1968). Moffett finds his discourse model in a Piagetian theory of intellectual development which conceptualizes cognitive development as corresponding with the growth of increasing degrees of abstraction. Moffett says:

The concept that I believe will most likely permit us to think at once about mental development and the structure of discourse is the concept of abstraction which can apply equally well to thought and to language (p. 18).

Thus, Moffett proposes a naturally structured theory of discourse which analyzes the subject into a progression of speaker-audience relationships, subjects, logical sequence, and literary form. His most recent presentation of his schema appears in Active Voice: A Writing Program Across the Curriculum (1981). The main lines of progression are as follows:

## Schema of Discourse

(reading down)

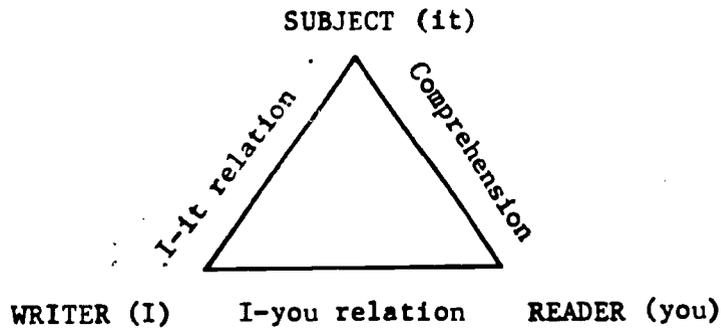
### Progression of Speaker-Audience Relationship

Thinking to oneself	Inner verbalization
Speaking to another person face to face	Outer vocalization
Writing to a known party	Informal writing
Writing to a mass, anonymous audience	Publication

### Progression of Speaker-Subject Relationship

Recording what is happening	Drama	The chronologic of on-going perceptual selection
Reporting what happened	Narrative	The chronologic of memory selection
Generalizing what happens	Exposition	The analogic of class inclusion and exclusion
Inferring what will, may, or could be true	Logical Argumentation	The tautologic of transformation and combination

On one dimension, Moffett displays the kinds of discourse in a continuum of time abstractions: drama records what is happening, narrative reports what happened, exposition generalizes about what happens, and argument theorizes about what may happen. The temporal movement is from present to past to future. On a second dimension Moffett orders discourse in a continuum of distance (space) between participants, moving in abstractive range from I-you in relation to readers and from I-it in relation to the subject. Moffett symbolizes this latter relationship as follows (Moffett and Wagner, 1976, p. 15):



The audience (reader) relation is a movement from self through introspection and reverie through an immediate audience to a mass audience removed in time and space. The relation to subject moves from personal, subjective response to impersonal, objective analysis.

Writing assignments that conform to Moffett's theory of discourse ask the student to:

...draw subjects from actual personal observation and to abstract this material in ways that entail increasingly sophisticated and artful decisions; assume a more and more remote audience; lead from vernacular style to literary style, from improvisation to composition; and open up for the student progressively higher realms of abstraction (Moffett, 1981, p. 6).

James Britton. At the University of London (1966-71) James Britton and his associates have developed a comprehensive multi-dimensional model of discourse derived during a five-year Schools Council Project on the Written Language of 11-18 Year-Olds. This British research team found the traditional discourse categories from the "predominantly Scottish tradition of rhetoric" wanting in a number of respects:

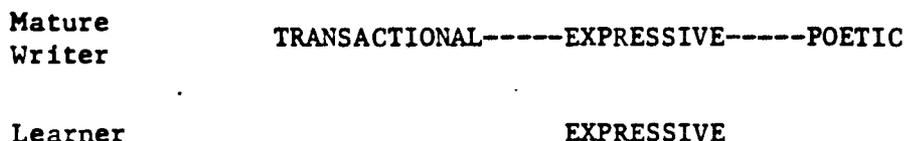
...they are derived from an examination of the finished products of professional writers, from whose work come both the categories and the rules for producing instances of them. The tradition is profoundly prescriptive and shows little inclination to observe the writing process: its concern is with how people should write rather than how they do. It can scarcely, therefore, be helpful in studying the emergence of mature writers from young writers (Britton, 1975, p. 4).

So instead of following the traditional categories, Britton's research team developed a two-dimensional model which on one plane classifies writing according to the predominant function that it performs. On another plane it addresses the audience dimension in somewhat the same manner as does Moffett's scheme.

Britton's audience scale moves from the self to the teacher to an unknown audience. An innovative elaboration of concept of teacher as audience appears among the distinctions portrayed below on the scale (pp. 116-129):

SELF	OTHER			
I	TEACHER	WIDER (KNOWN)	UNKNOWN	YOU
Trusted Adult	Gen'l Edu- cated Adult	Particular Educated Profes- sional	Examiner	

The main categories in Britton's scheme of the functions of written utterances are shown in diagrammatic form below as they are presented in his report (p. 83):



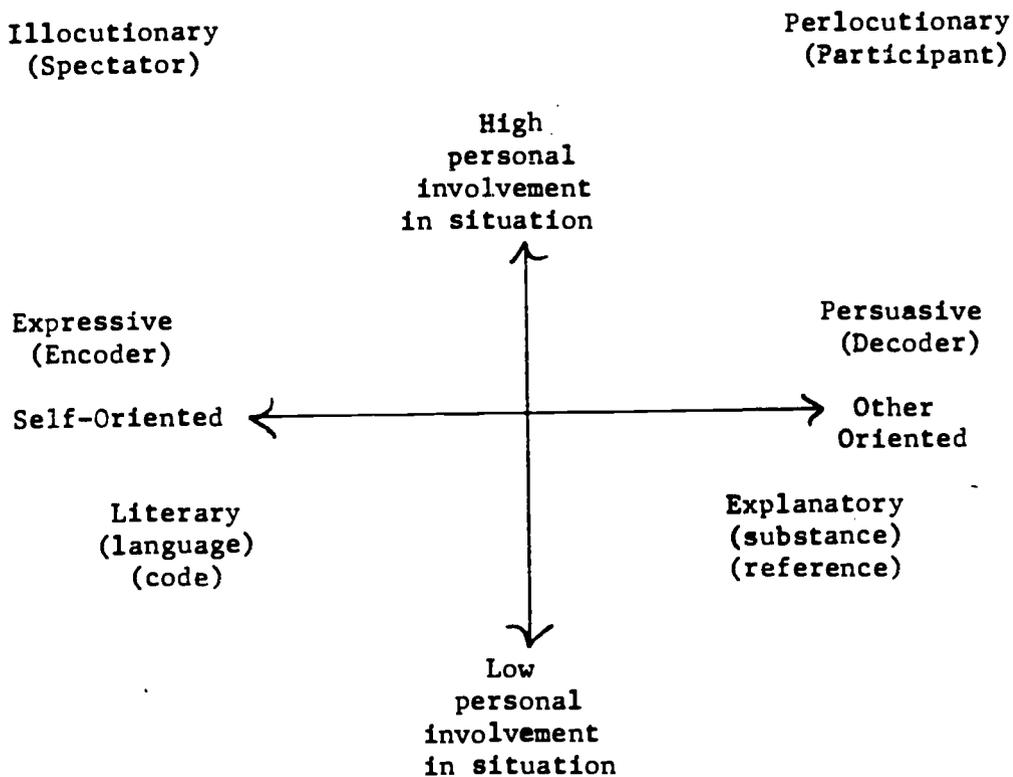
The expressive form of writing is seen as "a kind of matrix from which different forms of mature writing are developed" (p. 83). The expressive form is defined as "language close to the self, revealing the speaker, verbalizing his consciousness, displaying his close relationship with the reader." Transactional form is defined as "language to get things done, i.e., it is concerned with an end outside itself." The poetic form is defined as "a verbal construct, patterned verbalization of the writer's feeling and ideas. This category is not restricted to poems but would include such writings as a short story, a play, a shaped autobiographical episode." Neither the expressive nor the poetic forms are further subdivided, but the transactional is divided into informative and instructive (conative), and each of these has further subdivisions. The instructive (conative) is divided into the regulative and the persuasive functions. The informative is divided into categories following Moffett's categories of record, report, generalized narrative or descriptive information, analogic with a low level of generalization, through higher levels of generalization and speculative (tautological) forms (pp. 88-105). The informative category at the lowest level records concrete experience and moves progressively to generalize, abstract, speculate, and theorize at the highest levels.

Richard Lloyd-Jones and Carl Klaus. Since both assessment and instruction are based on explicit or assumed models of the nature and purpose of discourse, it is important to understand the main components of several of the leading models that now dominate

practice. Thus, we have examined the traditional models which classify products of writing and the leading contemporary models which describe processes and functions, as well as/in addition to products of writing. Lloyd-Jones makes a similar point in describing how he and Carl Klaus selected a discourse model to guide the procedures of the National Assessment of Educational Progress:

In order to report precisely how people manage different types of discourse, one must have a model of discourse which permits the identification of limited types of discourse and the creation of exercises which stimulate writing in the appropriate range but not beyond it (p. 37).

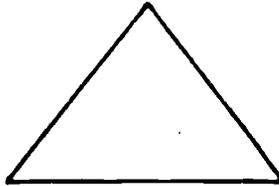
Consequently Lloyd-Jones and Klaus selected a three-part model after rejecting several others as being either too simple or too complicated. For example, the model below was rejected as too complicated for the purposes of the national assessment (p. 40). Even so it remained a useful explanatory model.



Although they acknowledge that usually there is a blend of purposes and associated forms in actual discourse, Lloyd-Jones and Klaus settled on a tri-polar scheme from which they expected to derive the primary rhetorical traits associated with each type of writing represented. The model they finally developed follows:

Explanatory Discourse  
(Subject oriented)

Expressive Discourse  
(Discourser oriented)



Persuasive Discourse  
(Audience oriented)

It is this simple model that has guided the design of exercises for NAEP since 1974.

Skills Model Limitations. The discourse models of Kinneavy, Moffett, Britton, and Lloyd-Jones and Klaus all stand in contrast to the skills model of writing which treats the subject as a series of discrete activities or skills, each to be practiced and mastered in turn. Kinneavy says:

all four of these authors--Moffett, Britton, D'Angelo [and himself]--believe that composition is best taught with examples of full discourse. The concerns with mechanics, sentence structure, style, even invention and arrangement are best seen in the act of handling a full discourse. Moffett (1968, pp. 205-206) decries what he calls the particle approach... (1980, p. 38).

Kinneavy continues with citations from each of the other authors to show their preference for holistic approaches over the skills model. It should be evident that assumptions about the nature of writing and models of the process have consequences for the design of assessment instruments.

Although the discourse theories of Kinneavy, Moffett, and Britton have been tested in classrooms, none except Britton's has received extensive trial under research conditions. It is time now to look more closely at the research contribution to the design of prompts for assessment of writing.

### Research as a Source of Knowledge

Basic Research Questions. As we have seen there are a good many assumptions in practice and theory about requisite features in a writing instruction assignment or assessment prompt. But relatively few details are yet available from research against which one might check these assumptions when one is designing prompts for writing assessments. Various researchers in the field of writing have noted this particular gap in available research information (Odell, Cooper, and Courts, 1978; Odell, 1979; Greenberg, 1981 Hoetker, 1981). The kinds of questions that are being asked indicate the current areas of need in the study of the means used to elicit writing for purposes

of assessment:

Questions raised by Odell, Cooper, and Courts (1978, p. 11):

How should researchers frame a writing task so as to obtain the best possible work from students?

Must researchers, as Sanders and Littlefield (1975) claim, provide a full rhetorical context, that is, information about speaker, subject, audience, and purpose?

Is there any aspect of the rhetorical context that we need not include in a writing task? Would an assignment that, for example, specified speaker, subject, and audience but not purpose, elicit writing that differed significantly from writing prompted by an assignment that specified a full rhetorical context?

Questions raised by Odell (1979, p. 41):

Is it in fact true that different kinds of writing tasks elicit different kinds of writing performance from students?

Are there some kinds of tasks in which purpose seems a more important consideration than it does in other kinds of tasks?

Does one writing task elicit a greater number of abstract (or connotative or formal) word choices than do other tasks?

Do different writing tasks lead students to use, on average, longer T-units (or more final free modifiers or more adjective modifiers...) than do other tasks?

Do different tasks lead students to use different types of transitional relationships or to use paragraphs that fill different types of functions?

Questions raised by Greenberg (1981, p. 8):

How do students read writing tasks? Which aspects of the directions do they understand? Or use? Or ignore?

How do students interpret writing tasks? What kinds of details do they think that the task is asking for?

How do students react to writing tasks? What factors in a task cause a student to perceive it as "easy" or "difficult?"

How are students' responses to writing tasks influenced by situational factors (test vs. nontest context, in-class vs. out-of-class assignment, timed vs. untimed composition, student-initiated vs. teacher-initiated tasks, and so forth)?

Literature Search. A recent search of the literature conducted at the beginning of this study in fall 1981 confirms that there are few answers to the kinds of questions that researchers are asking about the nature of the effects of writing prompts upon performance. For this study, the standard data bases were searched (ERIC, Dissertation Abstracts, Psychology Abstracts, Social Sciences Index, Language and Language Behavior Abstracts Index, etc.). Using such descriptors as 1) writing topic effect, 2) writing topic analysis, 3) writing evaluation criteria, 4) writing evaluation methods, 5) writing in relation to assessment, etc., some ninety abstracts were thus retrieved and followed up. But this search proved disappointing with respect to the number of items dealing directly with properties of writing tasks and their effects upon writers. Only one item in this set, a study of the effects of writing for different audiences in three different modes (Crowhurst, 1978), could be considered to reflect a research interest in the differential effects of writing stimuli. Nine other items offered bits of information about writing prompts, embedded in the context of larger research interests in evaluating writing.

Inadequate Research Report Practices. Unfortunately, current practice in reporting research results tends to gloss over or ignore topic effects in the very studies of writing where topics have initiated the written products under investigation. When topic effects are not the focus of experimental interest in a study, the information reported about the topic--its actual text and the presentation of any effects noted--tends to be cursory and oblique. Even such a notable and widely cited study as the Godshalk, et al (1966) investigation of the validity of the multiple-choice response mode in the measurement of writing neglected to quote the actual texts of the five criterion exercises used. Godshalk and his associates followed the rather common practice of describing rather than presenting the exact wording of the topic(s) used in investigations of writing. This reporting practice effectively disallows a reader of the study to make an independent judgment about the experimenter's classification of the writing tasks and to make an independent appraisal of explicit and implicit task demands. The consequences of this practice and its implications for retarding the accumulation of data about topic effects are examined in relation to several studies where topic effects were noted during the course of an investigation into other factors.

Nold and Freedman (1977) are exceptional, first for their attempt to control for topic effects in their carefully planned experimental study of factors that influence readers' judgments of essays, and second, for the level of detail which they devote to presenting

their essay questions. In designing their study, Nold and Freedman took precautions to insure that their results would not be topic specific by preparing two sets of two questions each, parallel in design and difficulty and requiring writing in only one mode (argumentative). Since Nold and Freedman include the texts of their questions, it is possible to inspect them and make independent judgments about their claims for equivalence in design and difficulty. And these texts do become a matter of interest when they report, "The kind of essay affected the amount of writing produced...." The two matched personal opinion topics elicited more writing and greater variation in response than did the two matched quotation comparison questions which elicited less writing and less variation (p. 170). Later in their discussion of the correlational findings, Nold and Freedman report that their study corroborates earlier ones that found length to be a predictor of quality: "The simple correlation between length and quality in our sample was  $-.57$ , indicating a high positive correlation. Our study shows that it is more damning to write a short essay than elevating to write a long one" (p. 173). Although this study does present considerable evidence of topic effects upon essay length and consequently upon its quality rating in a holistic scoring, this particular aspect of the study is never directly discussed in the report.

In a study conducted under the auspices of the Center for the Study of Evaluation at the University of California, Los Angeles, Winters (1980) set out to compare the differential effects of three types of scoring systems (impressionistic, analytic, and frequency counts) on a sample collected for this purpose. The report indicates that subjects were given 50 minutes to produce a 200 word expository essay on two test occasions, one week apart. "Two parallel topics were randomly assigned to subjects and counter-balanced to control for test occasion (see Appendix)" (Winters, 1980, p. 7). The texts of the writing tasks are quoted in full in this appendix. Winters's comments in her discussion of the findings on topic effects take up only two sentences: "Finally, the topics used in this study, which appear to have produced some unusual effects on writing performance were expository. There is no evidence that other modes of discourse or even other genres of expository topics would have produced the same results" (p. 19). Although this condition is cited as a limitation upon the generalizability of the study's findings, there is no explanation of what the "unusual effects" might have been. One must turn to another study from U.C.L.A. which used the same writing tasks for illumination of the "unusual effects."

Smith (1980) sought to compare the results of using different measuring strategies in assessing high school students' expository writing. This study of the relative merits of analytic scales, impressionistic ratings, and objective tests, consisted of two major components: the writing topics (and directions) and two forms of rating criteria. According to Smith, "The topics, designed to elicit like-samples of students writing, were intended to promote writing within the discourse domain of exposition...Task attributes guiding development of the topics included discourse mode, rhetorical purpose, content limits, and intended audience" (p. 6). The texts of

Smith's two topics are provided in an appendix and are identical to the two used by Winters. Because Smith's study found a pattern of strong relationships across analytic subscales, she acknowledged that it would be "tempting to infer that the Analytic scale actually tapped a single unitary dimension of writing" (Smith, 1980, p. 10). She cautioned, however, against making such an interpretation because of "an important facet of the writing task which may have affected the results--the writing topic and directions" (p. 10). Smith suggests the possibility that the "uniformity in the rhetorical structures of the majority of the students' writing samples" may have resulted from the structural promptings provided in the directions accompanying the topics. Ultimately, Smith raises the question: "Was the relative lack of independence among the Analytic subscales--and the high correlation with Impressionistic scores--a function of the homogeneity of student responses" (p. 11)? An independent inspection of the two topics suggests the likelihood that there could well have been the interaction between topic, essay, and rating as indicated in Smith's discussion.

Another study from U.C.L.A.'s Center for the Study of Evaluation is concerned with the validation of alternate response modes for writing assessment (essays, paragraphs, and "selected response format"). Capell and Quellmalz (1980) sum up the whole writing task in just a single sentence: "Both narrative and expository writing samples were elicited, with essays being on the topics of drugs and violence, and paragraphs on the topic of alcohol use" (p. 3). And that is all that this study says about the writing stimuli used in this experiment. None of the texts of the instruments used in this study is provided in the report.

One does not have to read very many research reports about writing assessment to determine that there is no uniformity in the practice of providing the texts of writing tasks and other relevant details. The importance of including full information about the writing task variable cannot be overstated. There is a twofold loss when insufficient information is provided: It becomes impossible to make an independent judgment about the claims the researcher makes about the writing task characteristics, and it also becomes difficult to build a cumulative knowledge base about writing task properties and related effects. For example, if Kincaid (1953) had not included his writing assignments in his study of factors affecting the quality of student writing, Braddock, et al (1963) could not have later reinterpreted Kincaid's findings. Whereas Kincaid had concluded that the writing performance of poor writers varied significantly according to the topic assigned, Braddock and his associates read the results differently. They noted that actually the content of the three topics was very similar, but they did differ in the modes of discourse they called for. Thus, Braddock concluded that the Kincaid study suggested "that variation of the assignment from expository to argumentative mode of discourse did not seem to affect the average quality of the writing of a group of freshmen who were better writers as much as it did a group who were worse writers" (p. 8). On this basis, Braddock argued that

until more could be known about the effect of mode on writing performance, it seemed necessary to control for this element in planning assignments to be used for research on writing.

One finds frequent reference to the "classic" studies of the reliability of the grading of essay tests in English, mathematics, and history conducted by Starch and Elliott (1912; 1913; 1913). For example, this work is cited by Coffman (1971), Foley (1971), and Sax (1974). Sax even quotes Starch and Elliott's famous assertion about the unreliability of essay test ratings: "It is almost shocking to a mind with more than ordinary exactness to find that the range of marks given by different teachers to the same paper may be as large as 35 to 40 points" (Sax, p. 119). Unfortunately, from the way these findings are often cited, one might easily get the impression that Starch and Elliott's "paper" was an essay on a single topic. Actually the two "papers" used in the English part of this study were not essays. They had been written to answer a set of six discrete questions, among which were the following: "2) Give five requirements to be observed in the structure of a paragraph; 3) Write a business letter; 4) Define narration, coherence, unity; classify sentences rhetorically and grammatically. Illustrate or define. 5) Name all the masterpieces studied this year and name the author of each" (from Starch and Elliott as reprinted in Payne and McMorris, 1967, p. 55). What is truly shocking is that such a crudely conceived study is still being cited as a credible study of reliability to bolster claims about the "disadvantages and limitations of essay tests" (cf. Sax, 1974, p. 117). If the actual text of this "essay test" had been omitted from the Starch and Elliott report, it would be impossible now to see how misleading their research findings are. Their so-called essay test is obviously a test of knowledge of discrete items.

One can only wonder about the nature of the other early studies of reliability which contributed to a growing lack of confidence in essay examinations during the 1930's and 1940's and that ultimately led to their virtual abandonment for the next three decades in favor of the more reliably scored "objective" test.

Now that we have discussed the practical problem of using research as a source of knowledge in designing prompts for assessing writing, we will turn to consider evidence regarding the following elements: 1) selection of the subject and problems of prior knowledge, 2) wording of the question and the problem of comprehension, 3) selection of mode, 4) specification of rhetorical context--role, audience, purpose, and 5) contextual influences.

1. Selection of the subject and problems of prior knowledge. Various teachers, evaluators, and researchers express awareness of the need for judicious selection of the subject or topic to be used in an assessment or research experiment (Lloyd-Jones, 1977, p. 44; Harpin, 1976, p. 92-93; Hoetker, 1981, p. 3; Greenberg, 1981, p. 3ff).

William Harpin (1976, p. 93), in reporting the results of a

longitudinal study in junior school (ages 7.0 - 9.11) involving 300 students and 50 teachers in nine schools in England, makes the point:

A theory of function and audience, no matter how elegant, is incomplete without reference to the experience that is to be worked on and aimed at a reader. Many teachers and certainly the majority of those who collaborated with us in producing the writing, see choice of subject or starting point as the most formidable challenge, dominating all other concerns.

Harpin's appraisal of writing assignments is based upon observations during the normal course of classroom instruction in the schools participating in his study. Although his chapter, "Contexts for Writing" (pp. 91-110), deals with writing in classroom instruction contexts, it still has relevance for assessment purposes. Harpin found a broad range of "writing situations" and he classified 22 kinds of experiences used to stimulate writing. According to Harpin, the nature of the stimulus offered is best described as "an opposition between open and closed or convergent and divergent situations" (p. 94). He concludes that the same experience may give rise to very different writing outcomes, according to the focus of attention suggested" (p. 98). A single stimulus may be presented as an open or closed prompt. Harpin cites an example where two teachers in different project schools and working independently used the identical stimulus, a picture of fire-fighting in Jacobean times, as a starting-point for writing. Each teacher held a brief discussion prior to having the students write. One teacher asked his class to "write a story" about the picture and received a set of historical fiction papers, all except one in the third person. During the pre-writing discussion, the other teacher considered different ways the picture might be responded to and received the following range of responses: 1) the historical fiction form, but also an "eye-witness" or "participant" report involving a change from third to first person narrative; 2) an itemizing of the content of the picture; 3) a broadly aesthetic treatment ("all the faces look alike and the colors are very dull"); and 4) consideration of the historical significance (p. 94).

Harpin's evaluation of photographs or pictures as stimuli for writing is instructive. Apparently equivalent images "work fitfully or not at all." From the evidence of experiments and experience in the project, Harpin concluded that "one significant feature is the extent to which interpretation is 'closed' or 'open.' There appears to be a necessary level of ambiguity for a photograph to succeed in arousing uniquely individual associations. Too much ambiguity and nothing happens; too little and the responses show little or no variation" (p. 106). For the seven- to ten-year-olds, human interest seemed "if not essential then desirable" (p. 106).

Harpin's discussion and classification of writing tasks (as classroom assignments) (1976, pp. 38-44) is by all odds the most comprehensive treatment found during this literature review. He

lists the following possible ways of classifying writing tasks: 1) by content, subject matter (what it is about, e.g. the Dewey Decimal System of classification); 2) by form (poetry, prose, essay, report, etc.); 3) by audience (learned, popular, the distinctions of rhetoric); 4) by writer-audience relationship (Joos's "five clocks" with its account of the interrelation of language, style, and speech event and Moffett's spectrum of distance discussed above); 5) by writer and task (creative, free, intensive, imaginative, practical, factual, recording); 6) by function, purpose, intention (expressive, poetic, transactional as in Britton above; or to tell a story, to describe a thing, to produce an emotional effect as in Herbert Read and Bonamy Dobree; or the order of phenomena in space in descriptive writing, the order of phenomena in time in historical writing, real or fictitious; the order of thoughts in the mind in expository writing as in H.J.C. Grierson). Ultimately, Harpin concludes that an almost endless series of classifications is possible: "Writing will refuse to fit into a cut and dried scheme of labels, no matter how cunningly or patiently constructed" (p. 37). What is important, Harpin thinks, is to select the scheme that is of greatest value in teaching, and we might add, in assessment. Harpin's work has been discussed here at some length because it is not likely to be readily accessible in the United States.

The discovery of fitting subjects for assessment--ones that are neither too large, too insignificant, nor too abstract--continues to be an art rather than a science because there appears to be little more than anecdotal information available about which subjects "work" and which do not. And even the anecdotal information remains scattered and unanalyzed. Characteristic of this type of data is the National Assessment of Educational Progress experience with certain subjects as reported by Lloyd-Jones (1977, p. 44), as follows:

Certain images create trouble. "Bananas," for example, seems to provoke pornographic or scatological responses. Bill collectors seem to be exceedingly threatening to some people. When we tried an explanatory-persuasive exercise to evoke a serious letter to correct an error in computer billing, we got a number of amusing responses, but when we tried to revise the exercise to encourage humorous literary-expressive responses by making the situation more absurd, we found that the increase in the number who could joke about computers and overdue bills was small and, strangely, the number which dealt with the problem in serious, highly conventional terms became larger. In both versions a substantial number of respondents were merely hostile. (emphasis added)

Lloyd-Jones generalizes that expressive exercises tend to be more enthusiastically received than transactional ones. He thinks this is the case because expressive writing is "to some extent its own reward." He points out that "Make-believe transactions present a mild contradiction of terms and that may tempt one to underestimate the skills in handling practical situations." Lloyd-Jones concludes, "...finding likely topics within the range of all the respondents

and challenging enough to produce the desirous efforts despite the lack of any 'payoff' remains a problem."

The reports of three studies done in Florida (Brossell, 1981; Hoetker, 1981; and Hoetker, Brossell and Ash, 1981) and yet to be published offer important data about several aspects of topics including the effects of the subject choice on holistic score. These studies were conducted by Hoetker and Brossell in conjunction with the development of a writing examination to be administered as a Teacher Certification Examination after 1980. Hoetker (1979) completed a very comprehensive review of the literature for the Florida Department of Education and then later prepared a condensation of the earlier work which is referenced here (Hoetker, 1981). Brossell (1981) conducted an experimental study of the effects of rhetorical specification in topics. And the findings of both of these studies were used in the development of a pool of topics for future administrations of the Teacher Certification Examination. This work in Florida and the work of Karen Greenberg of Hunter College (1981a, 1981b) are among a handful of studies in which the focus of interest is upon the topic or prompt and its effect upon writing performance in an assessment context.

With regard to subject effects, Hoetker (1981) identifies a study done by Rushton and Young (1974) which, Hoetker reports,

compared essays written by British sixth-form public school boys with essays written by factory workers of the same age. On academic topics the language from the two samples displayed the expected differences between "elaborated" and "restricted" linguistic codes. But when the two groups were asked to write on a technical subject, the linguistic disadvantage of the working class boys disappeared (p. 6).

The Brossell (1981) study of rhetorical specification was based on six topics at three information levels (to be described later in this report under the section on rhetorical specification) that he administered under simulated test conditions to undergraduate majors in education at the University of South Florida and at Florida State University. Brossell found that mean scores on the six subject matters (range 5.97 - 6.35) did not constitute a significant difference (p. 11). Although the statistical analysis did not reveal subject effects, Brossell and his associate Barbara Ash did note "strikingly superior" writing on one of the topics, whereas all the others elicited essays that they found to be "a desultory lot." On Topic 6 (level 2 version) on "violence in the schools," the writers addressed the topic more quickly, had more to say, and wrote papers that were superior as a group and received the highest mean score (7.05) (pp. 12-13).

Hoetker, Brossell, and Ash (1981) field tested 33 topics with a population of undergraduates in education to create a pool of validated topics for future administrations of the Florida Teacher Certificate Examination. Their procedures enabled them to determine the

"popularity of topics and they found that some topics were clearly more popular (more often chosen) than others. (The topics cannot be revealed because of the need to keep the pool secure for future administrations of the examination.) But these researchers report that "except in two extreme cases, popularity was quite unrelated to mean scores of essays written on the topics" (p. 10).

In addition, the students in this study had completed a questionnaire which asked them to explain why they had rejected the topic on which they did not write:

...the student comments made us realize that we had ignored another equally important dimension of the matter: the source of the essay's content. We had not taken into account that, even if all students are to write personal essays, it makes a difference whether the topic allows students to give their opinions on a public issue or demands that they introspect and report their inner lives (p. 10).

The researchers found that they could classify the topics into "public" and "private" categories and when so classified, mean scores were almost identical. Because student preferences are expressed strongly and often, they recommend "that student preferences be taken into account by choosing one public and one private topic as options for each form of the examination" (p. 11).

Hoetker, Brossell, and Ash did find three bad topics in the group of 33 piloted. These three topics--"American neglect of the urban environment, " "favorite gadgets," and an invitation to describe a "dream home"--were seldom chosen and produced low mean scores (5.2). Students found them "difficult, uninteresting, inappropriate, and requiring special knowledge (p. 12). The researchers' comments on the "dream home" topics are particularly illuminating: one of the researchers had had excellent results with a similar topic during classroom writing instruction when students had several days to work out their ideas. But on the test occasion (written in the 45 minutes allowed) every essay was disorganized:

Apparently, the students had fallen into daydreaming on paper, putting down each new idea as it occurred in a stream-of-consciousness fashion. That these students could not write well on this topic, which had been successfully used in writing classes, is one reminder that testing writing is not the same as teaching writing (p. 13).

Karen Greenberg (1981b) has provided information about nine questions field-tested at the City University of New York for the CUNY Writing Skills Assessment Test, a mandated essay test given to all entering freshmen since the fall of 1978 for the purpose of diagnosing writing ability and proper placement in writing courses (Greenberg, 1981a, p. 1). Greenberg provides a "popularity" index for these topics which give the percentages of native English speaking and ESL students who liked each question. At the time of this

report, it had not yet been established whether the preferred topics elicited better writing. Both populations gave the following topic the highest ranking (83 percent native speakers, 81 percent ESL):

It is the responsibility of parents, not teachers, to give children information about sex. Sex education is a private matter. It should not be taught in private or public schools.

Both populations gave the following topic the lowest rating (38 percent native speakers, 24 percent ESL):

The government is considering drafting young men into the army. Young women have been asking for equal rights and equal responsibilities. In all fairness, young women should be drafted into the army if young men are drafted.

Both groups give the same rank orders to all the other topics except one on which there is a substantial difference of opinion. Seventy-nine percent of native speakers liked the following question, but only 52 percent of ESL students liked it:

Some people work too hard and too much. In their efforts to succeed at school and at work, they often do not have time to build good relationships with their families and friends. This effort to succeed is not worth it.

Additional analyses are currently being made of the results of those field-tested prompts used in the CUNY tests. As yet they are unavailable.

Abraham Stahl (1977) set out to study developmental and ethnic differences in the writing of Jewish youth of European and Oriental heritage in Israel by collecting writing samples on the same subject at grades 2, 5, 8, and 11. The topic given was as follows:

My home: write a description of your home, its rooms, and their contents in such a way that someone who has not visited it can form an idea of the way it looks.

Stahl ran into unpredicted subject effects across the grade levels assessed. For example, many of the students of European heritage at grade 11 refused to cooperate because the subject seemed "too childish" to them.

It is easy to propose a number of maxims about the selection of subjects for assessment, but it is somewhat more difficult to be scientifically exacting in selecting the subject matter that we ask our students to give their heads and hearts to in an examination. We know that we must somehow find subjects, relevant and accessible, that students will meet with enthusiasm. But just now research provides only a few glimmers of how this goal might be accomplished.

## 2. The Wording of Prompts and the Problem of Comprehension.

Braddock (1976, p. 120) was perhaps the first to question the difference between a reading test and a multiple-choice test of written expression. It is only very recently that the multiple-choice test instrument itself has come to be recognized as a text with content and linguistic properties that presuppose certain reading capacities, whether it happens to be called a mathematics test, a writing test, an intelligence test, or other test. All of these tests, as well as essay tests, are first of all tests of reading comprehension. It is, of course, obvious that the writing task in an essay examination must be understood before it can be accomplished. Both Hoetker (1981) and Greenberg (1981b) comment on the reading problem inherent in the statement of the writing prompt. Greenberg, for example, in an ex post facto analysis of the results of a study which manipulated topic variables "discovered that many of the students who participated in my study did not perceive or did not understand the different demands of the experimental questions" (1981b, p. 8).

There appears to be very little research to guide test-writers in the wording of prompts that has been done through the direct study of the effects of different wordings of prompts upon student writers or teacher readers. What information there is tends to be anecdotal and is mentioned in passing in the context of a larger focus of research interest. For example, Harpin comments on the significance of wording changes:

where intention and outcome failed to match...teachers had to shoulder much of the responsibility. Substituting one word for another, though apparently a trivial change, may profoundly affect the way a child interprets the task--"Describe what you saw as you walked through the fog" gave very different results from "Describe what you felt as you walked through the fog." A similar effect is caused by the change from "What does this music make you think of" to "What does this music make you feel?" (p. 109).

Karen Greenberg (1982) comments in a letter on the problems of devising eight topics every semester in order to test about 30,000 entering students at CUNY. She says,

The opportunities for lexical and syntactic misinterpretations of test topics are numerous. For example, last semester, one of the topics asked about "vigilante groups that patrol our city's streets and subways." (We assumed that most students knew about groups like the 'Guardian Angels.')...we got responses about the 'villain groups,' the 'vigorous groups,' the 'Garden Angels,' and so on.

There is always a potential problem with presumptions about prior knowledge. The "tennis shoes" topic used by the NAEP and widely studied as a model could be subject to regional problems of

comprehension. James D. Atwater (1981, p. 13) reminds us that even such a seemingly prosaic term as "tennis shoes" might have to be changed to "sneakers" in some parts of the country, especially in urban areas. Ruth (1980, p. 214) notes that "underlying the assumed common language of questioner and respondent there may be variable coding and interpretation procedures that must be taken into account." He cites an example given as an illustration of inadequate pre-testing in the Michigan Survey Research Center's Interviewer's Manual as actually an example of variable linguistic coding: "In one SRC economic survey, a questioner asked, 'Do you think the government should control profits or not?' One respondent replied, 'Certainly not. Only Heaven should control prophets (1969, pp. 2-6)."

The whole problem of "response error," as the phenomenon is known in survey research, is not well-understood in writing research. That is to say, there is no theory or research in writing assessment to account for instrument errors which derive from the properties of the testing instrument, respondent errors which derive from the personal characteristics of the respondents—the student writers and the raters, or for contextual errors which arise from failures to understand the special conventions of testing contexts (e.g. the request to write a personal letter to display competence in a testing context requires a different performance than what is required in a naturally occurring communicative event).

In an effort to better understand the effects of wording essay questions one way or another, Greenberg (1981a) set up an experimental study which she hoped would answer questions about the "agree/disagree" format and the impersonal nature of the essay question format used in the City University of New York Writing Skills Assessment Test administered to entering freshmen in the system.

In a carefully controlled experiment, Greenberg (1981) studied the effects of varying the "cognitive" and "experiential" demands of essay questions on the writing of CUNY entering freshmen. For the purpose of the study she designed questions with "high cognitive demands" and "low cognitive demands." The HCD question called for an evaluation of an issue and involved determination and application of appropriate criteria of evidence as well as a capacity to distinguish among facts, opinions, and values. The question format called for the writer to "agree or disagree" and defend his or her position with logical reasons. The LCD question called for an interpretation of an issue, to relate issues rather than to judge them, and it provided generalizations and criteria for making judgments. This question format called for the writer to "discuss" ideas and provided a number of options for structuring the response.

Greenberg also sought to control the "experiential" demands of these essay questions, and thus she formulated questions with "high experiential demands" and "low experiential demands." The HED questions invited writers to relate personal experiences in response to solicitations cued with the second person personal pronoun "you." The LED questions encouraged students to respond in a less personal way with factual data and abstract generalizations.

The questions were presented in uniform formats consisting essentially of two parts:

1. A short passage (20-60 words) which introduced the subject to be written upon.
2. The instructions. The evaluative questions (high cognitive) asked the question: Do you agree or disagree with this statement? And they also requested supporting explanation and illustration in detail. The interpretive question (low cognitive) asked for a discussion of the opening statement, proposed at least three strategies for accomplishing the task, and requested reasons for the interpretations.

Efforts were made to control the rhetorical, propositional, and syntactic difficulty of the experimental questions in order to determine the effects of the two main variables, the cognitive and experiential demands. Care also was taken in the selection of a subject to avoid tasks that might be boring or threatening or outside the experience of the writer. Greenberg expresses her awareness of the confounding effects of controversial subjects such as "abortion" upon both writers and readers, and so she chose "education" as an area of shared personal experience that could best be shaped into the experimental questions. Sixteen questions were then piloted with four instructors and 29 students, and eventually eight questions in four sets of two each were determined to meet the specifications for the four conditions: 1) high cognitive/low experiential demands, 2) high cognitive/high experiential demands, 3) low cognitive/low experiential demands, and 4) low cognitive/high experiential demands.

Greenberg hypothesized that there would be a significant interaction between the cognitive and experiential demand levels of the essay questions with the low cognitive/high experiential demands eliciting higher overall quality of writing, as measured by holistic scores. Greenberg expected that the relative ease of the structured interpretive question (low cognitive) combined with the interest and accessibility factors of personal experience subjects (high experiential) would sustain better writing performances than the more complex, more remote subjects.

Greenberg's hypotheses were not confirmed: "None of the four types of questions elicited a substantially superior measure of rhetorical, syntactic, or lexical performance" (p. 58). The types of changes made in the content and structure of the questions used in this experiment did not have any statistically significant effect on the students' quality or fluency of performance as measured by holistic score or as determined by supplementary analyses such as scorings for syntactic complexity, sentence control errors, word form errors, and essay length. Interestingly, the personalized forms of questions which used the second person pronoun "you" as a cue to elicit personal experiences did not produce better writing or

even get a higher frequency of first person "I's" in response (p. 59). Greenberg speculates that this latter finding may have resulted from the students' belief that the first person pronoun is inappropriate in an essay produced for academic purposes. They also may have little faith in themselves as "reliable authorities" (p. 59). Greenberg considers, among other reasons for the results, the possibility that test anxiety may have constrained students from comprehending the instructions and taking full advantage of suggested strategies. She also thinks that the fact that so many of the students in the sample were "inexperienced" writers who were in or had been in remedial courses may have affected the way these students attended to variations in question wording (p. 60).

In a paper related to this study, Greenberg (1981b, p. 8) comments,

the most important thing I learned from my research is that the relationship between any writing task and students' writing performance in response to that task is profoundly influenced by the nature of the encompassing situation. The effects of writing task variables differ depending on the nature of the student and the nature of the setting.

In interviews with some of the students participating in her study, Greenberg found that "many had not even read the entire question--they simply read half-way through and then began writing the standard five-paragraph essay that their English teachers had always required of them..." (p. 8). It would seem that both in the case of the Brossell study described below and in the case of the Greenberg study, students were "reading" the context of the school testing situation more closely than they were reading the test instrument itself. They may have thus responded automatically with what Lanham (1979) calls "The School Style." The practical wisdom of meeting the demands of a test situation--writing for a supposedly traditionally oriented evaluator-reader and avoiding offense, grammatical or ideological--may itself beget the bland strings of unqualified assertions that so many observers find in the persuasive discourse of students at all levels (see next section for extended discussion).

Aside from the Greenberg study just presented and the Brossell (1981) study to be considered below, the larger theoretical issues of prompt design, of wording and information load, have yet to be addressed in writing research. There has been little interest in estimating the frequency, magnitude, or underlying nature of prompt-wording effects. For this reason, it is difficult to find prompt-wording experiments, and we have only scattered anecdotal information, reported as illustrative warnings, but not further developed theoretically or empirically. For the time being we will have to turn to survey research (Payne, 1951; Oppenheim, 1966; Gordon, 1980) and to reading assessment research which has recently moved in this direction (Meier, 1973; Kennedy, 1972; Fillmore, 1981).

3. Selection of Mode. Of the several variables in the expression of writing tasks, one that has received considerable research attention is the specification of mode. Since Braddock (1963) called attention to the neglect of control of the variable, mode of discourse, by people doing research in composition, a number of studies (Crowhurst, 1978; Crowhurst and Piche, 1979; Perron, 1977; Whale and Robinson, 1978; Rosen, 1969) have been conducted.

In a series of studies, Marion Crowhurst and associates have found that syntactic complexity is affected by writing task variables such as the specification of mode and audience (Audience will be taken up later in this review). When Crowhurst and Piche (1979) had sixth and tenth graders write in the modes of argument, description, and narration, they found significant differences in syntactic complexity between grades 6 and 10 in the modes of argument and description but not in narration. They also found that the difference in writing task produced a greater difference between students in the same grade than was evident in the comparison of the work of students four grade levels apart. That is, there was a greater difference in T-unit length between narratives and arguments at grade 10 than there was across all three modes in grades 6 and 10.

In another study (1978) Crowhurst found argument papers more syntactically complex than narrative papers at grades 6, 10, and 12. Since her findings are in agreement with Perron, 1976; Rosen, 1969; San Jose, 1972, Crowhurst (1978) concludes:

It may be regarded as well established by research that a) differences in syntactic complexity are associated with differences in mode of discourse, and b) narrative writing is generally less syntactically complex than argument (p. 8).

Harold Rosen's findings are in a Ph.D. thesis (1969) which, though not published, had been made available with the author's permission in summary form by John Pearce (1974) in a Schools Council Publication. Rosen sampled fifth year pupils in a variety of schools on the O-level Examination pattern (Ordinary Level in English Language in the General Certificate of Education) in response to a series of writing tasks. He found a clear difference in the difficulty of various types of writing. "Narrative/descriptive" writing is least difficult, "discussion/speculative" writing is the most difficult. Also he found that differences between an individual student's performance on different kinds of writing could be very great. The most satisfactory measures of linguistic maturity appear to be of limited validity and correlate more strongly with the kind of writing being attempted than with the overall competence of the writer. Fifth year pupils do not progress in writing during the year. As their perception of the "ideal" examination essay develops, their spontaneity and competence remain static or decline. One-word essay titles, instead of allowing for flexible, imaginative treatment, yield uniformly less good results than others. Rosen found the best writing was generated by open-ended literary or poetic stimuli which encouraged commitment without requiring self-exposure. Rosen found

that students in the fifth form were not yet prepared to meet the examiners' model of "good writing," discursive prose which adopts a detached, speculative treatment of a general or public issue. Rosen concluded that given the unknowns in the formulation of essay topics, the strongest possible reservations were in order with regard to the matter of single-sample testing and the absence of pre-testing (pp. 55-56).

In his review of the literature on topic effects, Hoetker (1981) reports on studies by Rosen and others which have found similar connections between mode of discourse and syntactic complexity. In addition to these, there are still several other sources that focus on the difficulty students have with writing argument or persuasive discourse (Cooper, 1979; Neilson, 1979; Miller, 1980). There seems to be no doubt that writing argument is difficult for students at all levels where it is assigned through early college levels. But Freedman and Pringle (1981) are the first to shed some light on possible causes of the difficulty. In a study of the capability of 12 and 13 year-old students in composing arguments as opposed to narratives, they found that these students were far more successful in realizing the conventional schema for story structure than for the structure of an argument (98 percent of the students could embody narrative structure in their stories but only 12.5 percent could realize the classical argument patterns). Freedman and Pringle, in discussing their findings, feel it is too simplistic to attribute this failure to create well-wrought arguments to "lack of organization" or "lack of a sense of form or structure" because these students demonstrate extraordinary deftness in organizing narrative material. Thus, they conclude that while students have multiple opportunities to internalize narrative schemas, no such parallel opportunity exists to inculcate argument patterns. Freedman and Pringle also discuss the developmental aspect of written argument and its requirement of cognitive maturation. To be able to write argument, students must have had sufficient exposure to the genre through reading and time to allow their maturation levels to enable them to cope with complexities of the form.

These findings about differences in the difficulty of modes would seem to have implications for the selecting of subjects and the preparation of accompanying instructions. Just how phases of cognitive, affective, and moral development need to be taken into account in the formulation of writing assessment tasks is dealt with at great length in Wilkinson's (1980) account of the work of the Crediton Project in Crediton, Devon (U.K.), 1978-80. Wilkinson and associates studied the development of the written language of children 7-14 on four planes--cognitive, affective, moral, and stylistic. Four compositions--narrative, autobiographical, explanatory, and argumentative--were requested from groups of children at ages 7, 10, and 13 in the context of their normal lessons. The same four subjects were given to each group and the results were compared according to four "models" for the analysis of writing--cognitive, affective, moral, and stylistic. The study attempts to demonstrate the need for establishing criteria to judge composition

in a developmental context. No simple summary of this work is possible, for the authors argue that qualities of thought, feeling, moral judgment and style all must enter into a process of continuing assessment conducted as an integral part of the day-to-day instruction.

4. Specification of Rhetorical Context--Role, Audience, Purpose. Of all principles espoused for designing the writing assignment, specification of a full rhetorical context is the one that seems to be most strongly established in the canons of practice. The lineage of this notion goes back to classical rhetoric, and it has found vivid restatement at various times in the history of writing instruction. P.B. Ballard (1939) gives an account of it in his chapter on "The Hartog Method," based upon the work of Sir Philip Hartog as it appeared in The Writing of English (1907). Ballard sets forth the first principle in Hartog's method:

...a pupil should never be required to write a composition without his having clearly in mind the audience for whom it is intended, and the object he has in view in writing it. This principle is flagrantly violated not only in school exercises but also in the essays set at public examinations. Current practice is neatly summed up by Sir Philip Hartog in the formula: "Write anything about something for anybody." Nobody does this in ordinary life. To quote Samuel Butler: "It takes two people to say a thing--a sayee as well as a sayer. The one is as essential to any true saying as the other." He might have added that something else is essential: a motive for saying it....Nobody writes for the mere fun of expressing himself on paper: he addresses himself to somebody whom he wants to influence; he wants to question him, to inform him, to persuade him, or to convince him. He has always at the back of his mind this possible reader or circle of readers--a friend, a coterie, or the world at large. Without this recipient and responsive factor the business of writing resembles an attempt to work a battery with only one terminal. Whatever else you may get, you will get no electricity (pp. 79-80).

Many professionals have recommended full specification of the rhetorical context, perhaps less eloquently but none the less emphatically than Ballard. Thus it is not surprising that Hoetker (1981) was convinced at the end of his original review of the literature that full rhetorical specification was desirable. Hence, he recommended in that original version that topics for Florida's Teacher Certification Examination "should take the form of scenarios simulating writing tasks likely to come up in the course of a teacher's work" (p. 11). But the results of Brossell's (1981) study of the "information load" of topics caused Hoetker to change his recommendation to favor briefer, less fully specified topics instead.

In 1980, Hoetker's colleague at Florida State University, Gordon Brossell, ran an experiment to test the hypothesis that essay

questions that specify a full rhetorical context are superior to less complete versions of the same questions. Brossell used six topics written on three information levels to generate writing samples to evaluate, by holistic scoring procedures, the effects of information level and subject (p. 3):

Level 1, low information load, presented the topic in a brief phrase. Decisions about audience, purpose, form, mode of discourse, voice were left to the writer with no guidance provided.

Level 2, moderate information load presented a topic containing a general introductory statement. It then posed a what or a why question asking for expression of personal opinion.

Level 3, high information load, presented a hypothetical situation and requested the expression of personal opinion as in level 2. But this version also specified role, audience, purpose and form or mode.

Topic 6 is quoted here in its three levels of specification.

Level 1: Violence in the schools.

Level 2: According to recent reports in the news media, there has been a marked increase in incidents of violence in public schools. Why, in your view, does such violence occur?

Level 3: You are a member of a local school council made up of teachers and citizens. A recent increase in incidents of violence in the schools has gotten widespread coverage in the local media. As a teacher, you are aware of the problem, though you have not been personally involved in an incident. At its next meeting, the council elects to take some action. It asks each member to draft a statement setting forth his or her view on why such violence occurs. The statements will be published in the local newspaper.

Write that statement expressing your own personal views on the causes of violence in the schools.

The other five topics dealt with 1) the "basics" of education, 2) teaching as a career, 3) discipline in the classroom, 4) the four-day work week, and 5) American wastefulness.

The six topics were randomized and administered to a sample of

360 undergraduate education majors at Florida State University and at the University of South Florida. A panel of three raters read and holistically rated the essays according to criteria developed for the examination. The score assigned was the sum of the three ratings (3-12). Inter-rater reliability was .828.

The statistical analysis was designed to determine the effect of topic, information load of topic, and length of essay on holistic score. Subject matter of topics did not make a significant difference, but information level effects were much stronger, and essay length was significantly correlated with score. Level 3 essays from fully specified topics received the lowest mean score and had the shortest mean length. Brossell is careful to note that "the differences between these means and those of the other two information levels are not large or statistically significant. "But (emphasis his) they are contrary to what those who assume the superiority of a full rhetorical specification would predict" (p. 12). In his discussion of the results, Brossell notes that level 1 requires the writer to get the subject in focus and find an organizational scheme while Level 2 offers a definite beginning focus and even something of an organizational strategy. Even though Level 3 provided a full rhetorical context, it did not provide direct aid for organizing or focusing the writing. Brossell reports that "many writers failed to get beyond a rehash of the information given," and it "often seemed to act as a hindrance rather than the facilitator it was meant to be" (p. 17).

The results of Brossell's study have led Hoetker (1981) to wonder about the reading burden of the longer scenario type of writing prompt. He points out that:

...the more information students are given--the more language they have to process--the greater seem to be the opportunities for creative misreadings and simple confusion. Second, the more information students are given, the more difficult they seem to find it to get beyond the language of the topic to discover their own language (p. 10).

Derek Rowntree (1977) provides an insight into what might be happening to a writer as the level of specification grows in a given essay question. As Rowntree makes the following set of questions progressively more specific, we can see that with each added constraint there is an increasing likelihood of eliciting more convergent responses with a corresponding loss of opportunity for divergent responses.

1. What aspects of the political system of modern Sweden seem to you most worthy of comment?
2. Comment on the political stability of modern Sweden.
3. Explain the political stability of modern Sweden.
4. Identify and discuss three factors that might help explain modern Sweden's political stability.

5. Identify and discuss three factors that might help explain the emergence of a stable political system in Sweden despite the massive social and economic changes engendered by processes of modernization.

Rowntree points out that each version progresses successively to restrict the student's scope of response. With this restriction comes a gain in uniformity of response and comparability of answers, but there is also a loss of information about the possible range of variability. Since the question and the criteria for answering are preformulated, it can no longer be a measure of whether the student could have formulated the question and decided the relevant criteria. Rowntree concludes:

...in an attempt to ensure comparability and make the grading more accurate, we may have ended up comparing and grading less significant abilities in the student. Macnamara's Fallacy, perhaps--making the measurable important rather than the important measurable (p. 156).

Flower and Hayes' (1979) process model of composing begins with the "task environment" ("the world outside the writer's skin") which consists of "the rhetorical situation" ("the specifications of topic and audience to which a writer must respond") and "the text which the writer has produced so far" (pp. 90-91). Flower and Hayes (1980, p. 40) discuss "the rhetorical situation" in terms of "the demands of the rhetorical problem":

Whatever writers choose to say must ultimately conform to the structures posed by their purpose in writing, their sense of audience, and their projected selves or imagined roles. In essence writing is also a speech act and therefore subject to all the constraints of any interpersonal performance.

Humes (1980, p. 5) says that "Flower and Hayes consider the problem to be the most important element at the beginning of composing because writers solve 'only the problem they represent to themselves' (p. 8)." Later Humes (p. 5) also mentions that Flower and Hayes consider discovering how "the rhetorical problem works and affects the writer 'is an important goal for research' (p. 8)." On the basis of their own studies, Flower and Hayes (1980, pp. 41-44) point out the strategies that writers employ to solve the problem of excessive constraints. They do such things as "throw a constraint away," "partition the problem," "draw on a routine or well-learned procedure," and "plan." As Flower and Hayes put it, "Writers inevitably set such priorities in the way they define their Rhetorical Problem (e.g., this is a letter to Aunt Tilly, so you can safely ignore run-on sentences and fragments; she won't mind)" (p. 42).

It is doubtless true that Aunt Tilly doesn't mind a few grammatical lapses, but the unforgiving evaluator does mind. Thus, the evaluator needs to be wary in specifying audience, for an injudicious specification may have syntactic consequences which affect an assessment

rating. Crowhurst (1978, p. 11) found that a tenth grade boy writing in the mode of argument produced a substantially longer mean T-unit length when writing for a teacher audience than he did when writing for a best friend audience. The matter of posing an audience is tricky, as Steinberg (1980) notes:

Posing reasonable audiences is not as simple as it sounds. On the one hand, the audience should be specific enough to govern what facts the writer uses and what the tone and thrust of the written statements should be. On the other, the audience should not be so exotic as to invite laughter or irritation. I can remember a period here at Carnegie-Mellon University when, in our zeal for teaching problem solving in writing courses, we posited such bizarre audiences that we annoyed our students instead of motivating them. In one assignment, for example, we asked students to write an explanation of the use of the toothbrush for a native of the Canadian Arctic region who had never seen one....In one class...it was clear that the students were moving from amusement to annoyance when one day about midsemester I came into the room at the beginning of the hour and saw on the board something like the following: "Write an explanation for a one-armed paper hanger who is allergic to paste about how he can paper this room while standing on one foot without harming the newly shellacked floor" (p. 166).

The complexities of specifying audience in writing assignments are analyzed in detail by Long (1980) and by Park (1982). Long argues that the traditional ways of specifying audience in writing exercises by listing observable physical or occupational characteristics (white male university administrator) rests on the false assumption that "people sharing certain superficial qualities are alike in all other respects" (p. 223). It is a practice that he thinks would be called "stereotyping" in any other context. Long, therefore, calls for a redefinition of audience and the application of literary theory to a refinement of the concept. He turns to Walter Ong's concept of "the audience as a fiction" for his new model. Long quotes Ong who says "the writer must construct in his imagination, clearly or vaguely, an audience cast in some sort of role--entertainment seekers, reflective sharers of experience,...and so on...A reader has to play the role in which the author has cast him, which seldom coincides with his role in the rest of life." This view of the audience as a created fiction requires the writer to posit a new set of questions. Long says,

Rather than beginning with the traditional question, "who is my audience?" we now begin with, "who do I want my audience to be?" Rather than encouraging a superficial, stereotyped view of the reader, we are asking the student to begin with a statement about the audience she wants to create...This leads directly to question of method: what distance between reader and subject should be established?

What of diction and the creation of tone? What pieces of information do I want the reader to take for granted?, etc. (p. 225).

Douglas Park (1982) offers a lengthy, astute analysis of the several meanings of "audience." He finds a number of conceptual traps in the way audience is typically used. When writers are told "Consider your audience," it is too simple to take this to mean concentrate upon some particular person or persons. Writers attend to a number of different kinds of issues when they think about audience. According to Park (p. 249) the meanings of audience tend to move in two directions: one movement is toward actual people external to the text; the other movement is toward the text itself and the audience implied there. But, however real the readers are outside the text, "the writer must create a context into which readers may enter and to varying degrees become the audience that is implied there" (p. 249).

Park finds the two general meanings of audience dividing into four more specific meanings (p. 250):

1. Anyone who happens to listen to or read a given discourse: "The audience applauded."
2. External readers or listeners as they are involved in the rhetorical situation: "The writer misjudged his audience."
3. The set of conceptions or awareness in the writer's consciousness that shape the discourse as something to be read or heard: "What audience do you have in mind?" represents the shorthand version of this set of awarenesses.
4. An ideal conception shadowed forth in the way the discourse itself defines and creates contexts for readers: "What does this paragraph suggest about audience?"

The latter two conceptions are most important for teachers or for persons interested in discourse.

After considerable elaboration upon these several meanings of audience, Park eventually considers the implications of his analysis for the teaching of audience. He believes teachers depend too heavily on the concrete image of audience as readers external to the text. He then comments on the strategy of using members of the class or the teacher as the audience for student papers:

The former strategy, of course, means that students write knowing that their papers will be "published" in the classroom. This practice has obvious powerful effects on how students see the act of writing, but it can be said to provide an audience only in the commonly used sense of external to listener or readers. Students' reading of one another's writing does not provide that crucial

ingredient, people rhetorically involved. The student writing for members of the class still has the problem of finding or inventing appropriate rhetorical contexts. In fact, useful as this strategy is, it may also create problems.

The awareness of specific critical readers, whether they be students or the teacher, may inhibit and complicate rather than simplify the problem of dealing with audience.

Park ultimately says, "the fact is that most of the time we want students to learn to write for a 'general' audience." That is to say we want them to write in relatively unstructured situations where little is given in the way of context and much remains to be invented by the writer." Park believes teachers need to be able to break audience problems down into specific issues and strategies. The strategies must "take shape against the background of the conventions appropriate for given kinds of writing. Much of the time it is not possible to separate a sense of audience from a sense of genre and convention....Probably writers [who] come to have an intimate sense of audience as convention [do so] by being readers of that kind of prose. I doubt if it can be taught very directly or very quickly" (p. 256). Ultimately, a clear understanding of audience must grow from a clear understanding of the kinds of discourse and their purpose in society.

It can be seen from these two analyses of the "meanings" of the term "audience" that it is a conceptually complex rhetorical construct. Thus, when researchers such as Scardemalia, Bracewell, and Bereiter (1978) suggest "that assignments that specify a particular audience in some detail act as important cues to writers because they help them 'decenter' from their content or information and adapt their knowledge to a reader" (quoted in Flower and Hayes, 1980, p. 45), one cannot be sure exactly what this finding means in terms of what was actually specified. There needs to be a good deal more critical analyses of different kinds of prose and more research into the composing process before we can be sure how best to guide test-makers in designing writing prompts.

Perhaps the writing task can be simplified conceptually, as Linda Flower proposes, into understanding the distinction between what she calls "writer-based" prose and "reader-based" prose. As she puts it, "Good writers know how to transform writer-based prose (which works well for them) into reader-based prose (which works for their readers as well)" (Flower, 1981, p. 144). Flower deals in the development of process strategies "to create a momentary common ground between the reader and the writer" (p. 122). In her view, then, the accomplishment of reader-based prose requires developing a critical awareness of features of the intended reader such as his knowledge, attitudes, and needs. It is necessary to go even beyond this base, however, for contemporary reading theory posits a "creative" reader active in the making of meaning, but a reader with certain limitations of short-term memory and certain expectations for the structural pattern of the text being read. Thus, the writer needs not only to know his reader, but he needs

to aid him by providing a framework or a context for new ideas (Flower, 1981, pp. 121-142, passim).

The evidence just recounted here from rhetorical theory (Long, 1980; Park, 1982) and the research into the cognitive processes of writing (Flower and Hayes, 1979; Humes, 1981; Flower, 1981) make moot the wisdom of declaring a particular role and audience for writers. To require a writer to assume the role of an advertising copy writer and prepare an advertising campaign for the executives of Acme Beer (topic idea from Sarah Freedman) presupposes prior knowledge on the part of the writer of the nature of the copywriting role, advertisement and report genre, and of Acme Beer executives as prospective readers of the advertising campaign material. An exceedingly complex task becomes even more convoluted as the writer realizes that in the real-world context of the assessment situation, his only real reader(s) will be the evaluator(s).

5. Contextual Influences on Assessment. We are just beginning to understand the nature of context as an encompassing framework around the process of writing and assessment. Staton (1981, p. vii), reminds us that "there is research to suggest that the greatest failure of writing instruction and assessment comes in not creating conditions under which students would have reason to compose a text (Graves, 1978, 1979)." Roger Shuy (1981) raises a number of questions about assessment and the constraining conditions imposed upon the writer. He concludes that

What we are assessing is not writers' ability to think, organize, sequence, explain, persuade, narrate, or describe, but rather their ability to do these things under the special set of circumstances created for the writing assessment (p. 170).

Shuy chooses an absurd metaphor to convey the idea that assessment measures only the ability to do a specific task under constraining conditions. Our approaches to writing assessment can be likened to an approach to the assessment of walking which assesses "the ability to walk on a slippery pavement with a broken toe and high-heeled shoes" (p. 171). We would not necessarily learn anything about ability to write or to walk in this approach to assessment, according to Shuy.

Michael Clark (1980) makes an astute analysis of the power relation that holds between participants in an assessment situation. To portray this condition he draws on the famous work of William Labov who examined the interviewing techniques of interviewers whose results led to classification of their subject as "linguistically deprived." He quotes two key passages from Labov which he feels speak equally well to the situation in writing assessment, although Labov's observations apply directly to an interview. First, Labov points out the origins of a subject's defensive maneuvers in an interview situation:

The child is in an asymmetrical situation where anything he says can literally be held against him. He has learned a number of devices to avoid saying anything in the situation,

and he works very hard to achieve this end....If one takes this interview as a measure of the verbal capacity of the child, it must be at his capacity to defend himself in a hostile and threatening situation (1969, p. 8) (quoted in Clark, 1980, p. 130).

In a second passage quoted by Clark, Labov provides the underlying interpretation of an interviewer's question such as "What do you think of capital punishment?"

the speaker's interpretation of these requests, and the action he believes appropriate in response is completely uncontrolled. One can view these test stimuli as requests for information, commands for action, as threats of punishment, or as meaningless sequences of words. They are probably intended as something altogether different: as requests for display (1969, p. 23) (quoted in Clark, 1980, p. 130).

Although the actual stimulus question may be the same for each respondent, virtually all of the other variables of the communication context may differ from respondent to respondent. Consequently, says Clark, "To judge all the samples elicited from the interviews by the same criteria is therefore to render the evaluations meaningless" (p. 131).

Clark declares that "any evaluation of speech or writing that ignores the extraverbal 'sociolinguistic' feature of the context in which the language was elicited is irrelevant as a measure of communicative competence. Furthermore, any assessment of language that purports to measure verbal capacity outside of a specific context--'Good Writing'--is therefore either deliberately misleading its users or is looking for something that is precluded by the test itself" (p. 131). After advancing this line of argument, Clark makes the claim that the situation in Freshman English and in any of the evaluative contexts is like that described above, and "shows our assignments and evaluations often proceed from the same false assumptions about the context-free nature of language acts" (p. 131). Clark then proceeds to describe an approach to assessment examinations of incoming students at the University of Michigan which he says is designed and read taking into account some of the insight summarized above. No examination question is provided as a sample, but the pattern is described. It would be misleading to summarize the examination procedures without providing Clark's interpretation of them, and so that will not be attempted here. But one does have to wonder to what extent the University of Michigan assessment procedure, though much refined, escapes the decontextualizing processes of other examinations. One also has to wonder how the students in the Michigan test escape "the role established for them by the context of the test" any more readily than the participants in any other test situation.

There is a confirmation of aspects of "sociolinguistic" variables of the assessment situation in Pianko's (1979) study of the processes of college freshman writers. Pianko finds a number of "interesting paradoxes" in the methods currently used for teaching writing.

In the first instance, students do not view writing which has the context specifically set by the teacher and which must be completed within the constraints of a class meeting as an activity that is worth committing themselves to. The limitations placed by the typical school writing activity negate the possibility for greater elaboration, commitment, and concern. Yet many instructors insist that the most effective way to evaluate students' writing abilities is to have the writing controlled for topic, place, and time. According to the students in this study, such a writing activity does not permit sufficient time for them to re-group their energies and thoughts; therefore, they merely attempt to complete the assignment in some expedient fashion and "give the teachers what they want." So, in fact, what writing teachers are actually evaluating is how well students follow instructions, not how well they write (p. 18).

Pianko's work appears to offer empirical validation of the sociolinguistic insights of Labov, Clark, and others mentioned above.

In 1979-80 Clark and Florio (1982) undertook a naturalistic study of schooling and the acquisition of literacy. They sought to document how two teachers and their respective students produce occasions for writing by means of interaction, to describe the process of teacher planning for writing instruction, to develop a typology of the diverse occasions for writing in each classroom, and to examine these occasions as contexts for writing that may make differing cognitive and social demands on the students as writers (p. 27). It is the latter purpose of this study which is of interest here. Clark and Florio identify four broad functions of writing in a second/third grade classroom and in a sixth grade classroom (p. 159):

1. writing to participate in community,
2. writing to know oneself and others,
3. writing to occupy free time, and
4. writing to demonstrate academic competence.

We will consider here only "writing to demonstrate academic competence," for we seem to have here a clear case of what Labov called the "display function" of an assessment. According to Clark and Florio, in classrooms, academic writing "is a teacher-initiated activity and it is the only type of writing that received formal evaluation from the teacher..."

It is the only type of writing that was typically both composed and formatted by an outside third party--the publisher. This fact, so commonplace in classrooms in

Clark and Florio conclude "that ambiguity about the audience of a student writing activity can affect both the interpretation of the task and the students' performance of the writing in diverse ways" (p. 67).

Further insight into the school as a context for writing is reflected in the Applebee (1980) report on writing in the secondary school. This study reports results from classroom observations of a year's writing assignments in a comprehensive city high school and in a university-associated laboratory school and from a questionnaire submitted to a national sample of teachers in six subject areas.

The observational study revealed that an average of 44 percent of observed lesson time was involved with writing activities. Mechanical uses of writing (short answer, fill-in-the-blank tests) occurred 24 percent of the time, note taking 17 percent, and writing of paragraph length or longer occurred 3 percent of the observed time. Homework assignments of at least a paragraph length occurred 3 percent of the time. The writing of paragraph length occurred most frequently in English classes, averaging 10 percent of lesson time. The most prevalent writing task was classified as requiring informational writing; imaginative uses were limited for the most part to English classes, but reported by less than half of the students.

The national survey of teachers in all subject areas indicated that they made frequent use of some writing-related activities but these were dominated by note taking and short-answer responses. Paragraph length writing was reported in 27 percent of classes at grade nine and in 36 percent at grade eleven. The writing samples supplied reflected informational uses of writing in 85 percent of the cases. The teacher in the role of examiner was the prime audience for student writing in all subject areas. Only 10 percent of the teachers reported that the writing was read by other students. Of the writing samples submitted by surveyed teachers, 88 percent were addressed to the teacher as the primary audience. Generally, 48 percent of the sample was informational writing addressed to the teacher as examiner.

The consequences of the emphasis on the more mechanical forms of writing required across the secondary curriculum are described by Applebee (p. 141):

Analyzed as writing activities, such tasks are characterized by a separation of the problem of constructing coherent text in a language appropriate to the subject areas from the problem of remembering subject area information and concepts. Essentially the teacher takes over all the difficulties inherent in using language appropriate to a subject area--including much of the specialized vocabulary and rules of procedure which are embedded in the text--and leaves the student only the task of mechanically "slotting-in" the missing information (p. 141).

...

our culture, was at first taken for granted by the researchers. However, when students were asked to sort and talk about their written work, they consistently grouped worksheets and workbook pages together because they were produced "by machine" (p. 165).

These academic performance writing activities engaged students in a variety of discourse functions including explanation and description, but these activities were not generated or ultimately controlled by the students:

the topic was constrained, the format of the writing was limited, and the function of the writing was pre-determined by others (p. 166).

Clark and Florio provide an interesting example of how a teacher-imposed task that is intended to stimulate private, free writing (fulfilling the writing-to-know-oneself function) is actually, in effect, interpreted by the students in a way unintended by the teacher and becomes, in effect, writing to display competence. One teacher set up an activity called "Diary Time" which was intended to be mainly private. Students were to be free of concerns about audience and form: they were to write their own thoughts. But as it worked out, Diary Time had some features that made it similar to more formal types of school writing. Clark and Florio describe the Diary Time writing situation (p. 65):

The facts that Diary Time was a teacher-imposed task, that the teacher knew and taught about "the rules" of diary keeping, that the teacher read aloud a part of her own first journal entry, and that the teacher collected and stored the diaries in her filing cabinet all supported the idea of the "teacher as audience." Student concerns about accurate spelling, questions about how the teacher "will correct them," and teacher assistance for those who could not think of what to write constitute additional evidence of teacher as audience.

Clark and Florio found this to be an instance of an assignment that sent a "mixed message" to the students causing confusion in their efforts to accomplish it. They observed three types of student response (p. 66-67):

1. Some treated the Diary Time as "just another school writing task" and wrote what they thought would please the teacher.
2. Some treated the Diary Time as a reflective writing task where the usual "rules" of school writing were suspended.
3. Some wrote nothing and waited for the teacher to help them think of what to write and thereby remove the ambiguity from this new type of writing task.

One of the major problems with an overemphasis on mechanical writing tasks is that the students may never learn to use such resources on their own, relying instead upon the structure or scaffold that the teacher has provided (p. 143).

...

Because the emphasis is on specific items of information, rather than on the way those items are integrated and presented in coherent prose, such writing situations provide little opportunity for instruction that might help students develop specific writing skills. For learning to write well, the most effective writing situation will be one in which the effectiveness of the writing matters--where the student can savor the success of having presented a convincing argument, or struggle with the problems of having failed to do so (pp. 143-144).

What emerges from the Applebee study that has relevance to the assessment condition is that students live in a classroom setting that offers little opportunity to practice the writing skills upon which they may be formally assessed. Much of the classroom writing that the students do is assigned in a test situation rather than in an instructional one. The test situation itself tends to be undemanding of integrative compositional processes because the dominant test forms require only the "slotting-in" of information or short answers which do not test one's mettle at structuring an idea. The informational writing that constitutes the principal type of assignment tends to require only relatively isolated bits of information, often only in the context of highly prestructured tasks (p. 142).

These teacher-as-examiner prestructured tasks would seem to constitute scant preparation for the writing assessment tasks frequently demanded. These teacher-prepared texts with gaps to be filled in seem to call for a different sort of cognitive ability than is required for creating and monitoring the development of one's own text in an impromptu text produced for a writing test (see Bartlett, 1981, p. 23). The teacher's own writing assignments and assessment tasks form an unobtrusive context for what later happens in the public examination, whether it occurs at the school level or in a national sample. Logic suggests that there should be preparatory work for what is to be assessed and that the examination itself should not constitute the only practice of particular task demands and discourse forms.

WHAT COUNTS AS RESEARCH?  
PROCEDURES OF THE STUDY

Knowledge Without Numbers

It has long been a requirement in educational research that one design studies which would generate statistics in order to formulate conclusions that the general community of educational researchers would consider binding. Grounded speculation which seemed to follow logically from evidence did not count as research. Common sense was suspect. But eventually some sociologists, cognitive scientists, and social theorists began to question our narrowly-conceived definitions of research. One of these sociologists stated the issue succinctly in relation to the study of human affairs: "...not everything that can be counted counts, and not everything that counts can be counted" (Cameron, 1963, p. 13). More recently, at the annual cognitive sciences conference held at the University of California, Berkeley in the fall of 1981, Robert Abelson of Yale's Artificial Intelligence Project, observed that

It is difficult to formalize aspects of human thought which are variable, disorderly, and seemingly irrational, or to build tightly principled models of realistic language processing in messy natural domains. (p. 1)

Abelson went on to characterize the clash of research traditions as being between the "neats" and the "scruffies" --in other words, between those "whose canons of hard science dictate a strategy of the isolation of idealized subsystems which can be modeled with elegant productive formalisms" and those other scientists who seek to deal with "the rough and tumble of life as it comes" (p. 1).

Something of this same type of ferment must have been evident in the September 1980 NIE-FIPSE Grantee Workshop that convened in California, for in his "Closing Remarks" Roger Shuy listed several of the currents of "uneasiness" that he had observed at the meeting. Among them was the uneasiness "about the way we have used data in the past. I think maybe the big battles in many disciplines are the battles over what constitutes good data....we still have a kind of uneasiness about data, how to use it, where it comes from, what it amounts to" (p. 166). Shuy went on to remark:

The old model of theory going to research going to practice is no longer viable. Rather, the view presented here was more like the view presented of writing--a recursive, iterative model, with theory, research and practice interacting (p. 168).

Meanwhile, in her "Welcome" at the same conference, Marcia Farr (Whiteman) addressed the question of methodology in writing research:

We are devising our own tools, choosing from among a variety of social-science disciplines and approaches as we go along. We must create structure out of what we find--out of data....Writing is a human activity; as such, what we really need to do in order to get valid information is pick our way along and eclectically use methodology from here and there, selecting techniques to help us learn what we want to know. Being eclectic, of course, does not preclude being rigorous in our thinking (p. 13).

And later in this same "Welcome" Marcia Farr added:

Often researchers get too far from the social contexts in which their work could be useful, and their work becomes so esoteric that it requires much further work to "translate" the findings into practice. It is also important for those who teach writing every day to participate in formulating the research questions...(p. 13).

The eclectic research strategies employed in the conduct of the various phases of this Writing Assessment Project find their authority in spokespersons from the research community such as those identified above as well as in those from the strongly articulated views of a group of British educationists (Stenhouse, 1981; Bassey, 1981; Guy, 1975; Volpe, 1981; Hogan, 1980; Elvin, 1975). It is not appropriate here to review the methodological issues addressed by all of these writers, but it does help the case being developed here to consider the position taken by Lawrence Stenhouse. First, Stenhouse (as quoted by Bassey, 1981, p. 74) suggests that

Our problem is to find approaches to research which produce theory which is of use both to practitioners of education and to practitioners of educational research and which enables both to act in the light of systematic intelligence.

Bassey continues to quote Stenhouse:

He goes on to assert that "the most important distinction in educational research at this moment is that between the study of samples and the study of cases"...it is clear that he expects both the study of samples and the study of cases to lead to generalizations.

Stenhouse distinguishes between predictive generalizations, which arise from the study of samples and are the form in which data are accumulated in

science, and retrospective generalizations, which can eventually arise from the analysis of case studies and are the form in which data are accumulated in history. He writes "while the predictive generalizations claim to supersede the need for individual judgment, retrospective generalizations seek to strengthen individual judgment where it cannot be superseded (p. 75).

It is this power of the "retrospective generalization" that is of particular interest in this study, for Stenhouse offers a way of making respectable the use of common sense and practical knowledge in educational communities. In a significant article, "What Counts as Research?" (1981), Stenhouse discusses the relevance of both science and history to practice, but it is his account of the significance of history-as-research method that is of prime importance in describing the rationale behind our procedures for consensually validating the report that constitutes Part II of this chapter.

Stenhouse discusses how history-as-research-method helps define the conditions for future action by summarizing experience and suggesting considerations to take into account as we make judgments about how to act. Thus, judgments of relevance about acting in any given case can be founded on "stock-taking," state-of-the-art reports, which Stenhouse conceives of as "contemporary histories" (p. 105). History summarizes the experience of action in such a way as "to strengthen judgment and revision of judgment in the planning of acts." Stenhouse further points out that the historian identifies issues just as often by ambiguity as by stating hypotheses. Further, the historian takes most of his terms from his subjects, thus "One great strength of history is that its vocabulary is accessible to those who are interested in the topic under discussion" (p. 108).

Stenhouse respects the social science experimental and analytic models, but he finds little in the reports of "mean results" that is of use in the fields of action in classrooms where teachers must make individual decisions. Two more quotations will suffice to define Stenhouse's position and advance the theoretical justification for several of the types of studies conducted in this project. Stenhouse summarizes:

In short, it seems that, while social science applied to education can produce results which help us to understand the ground rules of action, it cannot provide the basis for a technology of teaching which offers reliable guidance to the teacher. Predictions based upon statistical levels of confidence are applicable to action only when the same treatment must be given throughout the entire population. This condition does not apply in education. It is the

teacher's task to differentiate treatments.

...

The portrayal of cases offers to inform the judgment of actors--the administrators, teachers, pupils or parents--rather after the manner of history, by opening the research accounts to recognition and comparison and hence to criticism in the light of experience. Such a refinement of experienced practical judgment eludes the psycho-statistical model which strips the data of recognizable characteristics and content, and presents "findings" or "results," which are accessible to criticism only by replication or by technical attack on the design or conduct of the research (p. 107).

#### A Collaborative Model for Validating the "Theory of Idealized Practice"

So much of educational research deals with data that is so highly abstracted, so remote in the language and form of its expression, that the teachers and students who were the original creators of the data cannot recognize any human connection in the reported results. The results do not seem to relate to particular human beings interacting in particular contexts. Thus, we sought a method of investigation that might tap the store of wisdom possessed by teachers and test-makers practiced in the creation of writing prompts and the conduct of holistic writing assessment. Since we wanted a method that preserved the identity of experienced practical judgments about the design of writing prompts, we could not turn to any of the conventional statistical survey research models. We found no guidance in any extant work on educational research methodology, so, as Marcia Farr predicted might be necessary, we ended up inventing our own procedure for gathering our personal and professional knowledge and exposing it to critical scrutiny.

Catharine Keech as Informant. In effect, we used a teacher-researcher, Catharine Keech, as an informant to compose a document reflecting her perception of the essential procedures to follow in developing writing prompts for holistic assessment. Ms. Keech brought to her task the dual perspective of a classroom teacher and an experienced researcher in writing evaluation. Prior to entering graduate study at Berkeley, Ms. Keech had taught for twelve years, ranging from elementary grades through secondary school and college. In 1977, Ms. Keech went to work for the Evaluation Unit funded by the Carnegie Foundation to study the effectiveness of the Bay Area Writing Project. This Evaluation Unit, under the direction of Michael Scriven, assigned Ms. Keech as a research assistant to collect background information on how writing is evaluated in this country in order to approach the evaluation

of the Bay Area Writing Project. During the course of her investigation, Ms. Keech consulted Evans Alloway of Educational Testing Service, Joe Steele of the American College Testing Service, Rex Brown of the National Assessment of Educational Progress, Beth Brennehan of the California State Department of Education's Measurement Division, Jean Jensen of Las Lomas High School, Walnut Creek, California, Kate Blickhahn of Sir Francis Drake High School, Mill Valley among others. Then she met Allen Seder at the Educational Testing Service office in Berkeley, who invited her to read for the California High School Proficiency Examination. Later, under the auspices of this BAWP Writing Assessment Project, Ms. Keech spent a day with Gertrude Conlan and her staff at the Educational Testing Service office in Princeton learning about the procedures used there for developing and piloting writing prompts. From that meeting Ms. Keech brought back the Conlan checklist for writing prompts that is appended to Part II of this report. During her work for the Scriven Evaluation Unit, Ms. Keech also became acquainted with Edward White, the chief designer of writing examinations for admissions to the California State University system, and with Paul Diederich. During these years, Ms. Keech collected an extensive file of topics and developed an unusual depth of background knowledge of current practices in designing writing assessment prompts. Thus she came to this project prepared to write the draft of a document called, in its first version, "Topic Design: Views from the Profession."

The Review Panels. In the late fall of 1980, we sent a draft form of Catharine Keech's manuscript to two panels of reviewers. As project director, I wrote a cover letter asking each person to serve as a Critic-Reader of the paper provided. The letter mentioned that we desired to prepare a technical report on topic development from the point of view of teachers, researchers, and professional evaluators. (The people responding are identified in the report of the findings that follows this section.) The letter said, in part, "The report will accumulate illustrative 'good' and 'poor' topics along with sage advice on the problems of topic development and how to avoid them." We invited the readers to comment specifically on the "completeness, accuracy, organization, and style of the presentation" and to share with us any illuminating anecdotes from their own experience relating to "unexpected behavior of certain topics." We mentioned that we expected to incorporate into the final version of the report a summary of the comments received. That summary appears in the next section of this chapter.

The panel members had been selected on the basis of their many years of experience as readers in scoring sessions for holistic writing assessments, as developers of writing prompts themselves in school, state, and national assessments,

or as researchers interested in writing assessment. Among the panels were representatives from the Educational Testing Service, which pioneered the development of holistic assessment procedures, as well as several teachers who pioneered in adapting these procedures for use at local school or district levels.

#### A Recursive, Iterative Model: Theory, Research, and Practice Interacting

Roger Shuy, as mentioned above, noted the passing from favor of the old linear model of knowledge generation (theory-to-research-to-practice) in preference for a new recursive model of theory, research, and practice interacting. Though we make no claims for the perfection of our procedure, we think we have conducted a modest trial of this iterative model of a procedure for consensual validation of that professional wisdom which constitutes a "theory of idealized practice."

We created a medium for collection, exposure, and trial of ideas--the ideas and knowledge that arise from practical experience. Practical knowledge does not carry the authority of the processed knowledge of science in many quarters where naive interpretations and applications of "scientific method" to educational problems have led to an undervaluing of personal and professional knowledge. Personal knowledge is less orderly, less processed, less capable of generalization than the processed knowledge of science. But practical knowledge is rich in content; it is tied to specific contexts; it has been tested under fire. It is valuable for it has withstood tests of situational verification. The problem is that such knowledge remains in limited, private realms. It does not circulate, and therefore it does not profit by criticism from within the community it might serve. And this step is crucial.

This is why we submitted the original Keech account of practice to a circle of collegial critics. For when personal practical knowledge comes under the scrutiny of fellow professionals, that particularized body of ideas, experiences, hypotheses, theories, and practices comes into contention with the experiences of others in the community of practitioners. The limits of the validity of the ideas proposed are set through exchanges directed toward mutual learning. Though the community of professionals shares interests and commitments in problems of the measurement of writing, the process preserves the acceptability of disagreement (as is noted in the following report of results). Because different people have different ways of looking at and judging the processes of conducting writing assessment, a collective image of the reality of the process takes on a variety of perspectives more comprehensive than any individual contribution. Stenhouse (1980), in quite different language, makes many of these same points in answering the question: "What Counts as Research?" (pp. 108-112). He ends his quest with the observation: "Research is educational to the extent that it can be related to the practice of

education" (p. 113).

The next section reports the collaborative results of our efforts to refine our current knowledge about procedures for designing writing test prompts.

## RESULTS OF THE STUDY

The open-ended invitation to read and review Catharine Keech's preliminary version of "Practices in Designing Writing Test Prompts" elicited responses in a variety of forms: textual editing changes, marginal comments of varying length, and cover letters usually ranging from one to three pages. The frequency of marginal comments ranged from eight to sixty among those respondents choosing this mode of reaction. In one special case, the respondent provided twelve pages of comments, each enumerated and keyed to specific passages in the Keech text. The respondents offered general comments about the report as a whole, reactions to specific points raised in the report, relevant anecdotal information from personal experience in teaching and assessing writing, and various helpful suggestions for additions and changes that in many cases were incorporated into the final version appearing as Part II of this study. Most of the comments dealt with the idea content of the report, although a few dealt with format, organization, style. Many of the substantive comments reflected the respondent's own experience with the same or similar prompts.

The data provided through this form of panel review do not yield to quantitative treatment, yet identifiable themes and consensually validated recommendations for practice do emerge through the collation of responses from the community of professionals represented in this sample. The remainder of this report identifies the main themes and recommendations of the reviewers and provides illustrative quotations from their submissions. The findings are presented in two main sections: the first summarizes the responses of the writing instructors; the second summarizes the responses of the researcher/evaluators.

### Responses from the Panel of Writing Instructors

#### THE PANEL OF WRITING INSTRUCTORS

- \*1. Ruby Bernstein, English teacher, Northgate High School, Concord, California
- \*2. Mary Frances Claggett, English department chairman, Alameda High School, Alameda, California
3. Kim Davis, Subject A, University of California, Berkeley, California
4. Edmund Farrell, Professor of English Education, Department of Curriculum, University of Texas, Austin

\* Designates Bay Area Writing Project Teacher/Consultant

- \*5. Jean Jenson, English department chairman,  
Las Lomas High School, Walnut Creek, California
- \*6. Irving Peckham, English teacher, Live Oak High  
School, Morgan Hill, California
- 7. Bill Robinson, Director of Composition, English  
Department, San Francisco State University
- \*8. Bernard Tanner, English teacher, Palo Alto High  
School, Palo Alto, California

Value of the Study. The eight writing instructors who evaluated the trial version of the report were unanimous in their praise of the document in general. They all valued the attempt to compile information about the design of writing prompts and assured the researchers of the need for the guidance it aimed to offer. Ruby Bernstein, a high school teacher, thought that such a study would be especially significant for teachers of English and that it could be useful, as well, to teachers of subjects across the curriculum. She felt the work gave her insight into how "boring papers" produced by students might result from the way she had framed the initial topic, and she suggested workshops be conducted for teachers in the development of writing prompts.

Bill Robinson, Director of Composition at San Francisco State University, called the report a "very good job, covering in an intelligent and helpful way a really tough problem." Kim Davis, Director of Subject A at the University of California, Berkeley, said "This study should become required reading for all English teachers--a great many of whom give little hard thought, I suspect, to the writing tasks they continually set for students...." And Edmund Farrell, Professor of English Education at the University of Texas at Austin, though praising the efforts, suggested that the study should address itself more "to classroom teachers, who flounder in their attempts to create decent assignments." However, Farrell feared that the study won't get "the circulation it deserves," and suggested revising it for submission to NCTE/ERIC as a TRIP booklet. Kim Davis also mentioned widening the scope of the study to include creation of writing assignments in general--i.e., "routine classroom assignments"--not just for assessment purposes.

Classification of Writing Prompts by Discourse Mode. Most of the reviewers were enthusiastic about the draft of the report they received. They recognized it as a work in progress with great potential, and they welcomed this opportunity to offer constructive criticisms and to contribute to its fuller development. One feature of the report most in need of reconsideration, according to several of the reviewers, was the scheme adopted for classifying types of writing prompts. This scheme, based on James Britton's discourse model, sorted prompts into three broad categories: transactional, expressive, and

poetic. Both Bernard Tanner and Bill Robinson considered the weakness of this classification in some detail.

Bernard Tanner, high school English teacher, stated that he did not "buy Britton's intellectualized categorizing of youngsters' writings...." Tanner challenged Britton's functional classification of writing, explaining that he "could never be sure whether Britton's categories derived from the 'natural' modes adopted by children or whether they are the result, merely, of the kinds of tasks British teachers call for...." Tanner spoke of "the pinch of categorical definition" that overlooked too much. He also argued that "no categories prove exclusive of aspects of others," and he cited many examples of dubious classifications (in his view) to prove his point. Tanner himself finds the older rhetorical labels more helpful, and as he prepares a writing prompt, he asks himself the question: "What mix of modes of discourse do I wish to lead the writer toward?"

Another of the writing instructors, speaking from his vantage point at the college level, also considered the limitations of the classification scheme presented. Bill Robinson presented a strong argument against using Britton's functional triad, transactional - expressive - poetic:

I am not persuaded that the modern approaches to mode are any more helpful than Bain's, and Britton's are in fact heavily value-laden. Sending teachers to him, or worse, Emig for guidance is, I think, a mistake. In addition, the fact that you are addressing this paper to teachers or evaluators of children of wildly different ages and abilities seriously complicates your problem. Essential to Britton's categories are the various stances the writer assumes or is forced to assume vis-a-vis the audience. In his "expressive" mode, the audience is nonjudgmental to the point of being paternal or maternal. It seems to me highly unlikely that such an atmosphere can be set up in a testing situation with any but the teeny tiniest of little kids or the dopest. In addition, whether the mode is expressive or transactional, one still has to decide what form of writing one wishes to evaluate--or let us put it slightly differently--what skill one wishes the students to demonstrate. Does one wish to see whether they can tell a story of their own devising? (Note, however, that willy nilly one gets into testing creativity at this point.) Does one wish to see whether they can recount an anecdote from their experience--i.e., handle simple narrations? Does one wish to see whether they can discuss and evaluate an idea of some sort in terms of their own experience--i.e., handle simple exposition? Does one wish to see whether they can evaluate the pros and cons of a problem with which they can be expected

to be familiar--i.e., argue validly? I think it would be much more helpful to lay out the potential tasks to be tested in ways such as this rather than to examine some extremely dubious categories.

On the basis of these recommendations and others made by the researcher/evaluators, the final version of the report does not attempt to apply the Britton scheme to the classification of prompts for writing assessments.

The Content and Wording of Prompts. The teacher-reviewers offered many comments about the content and wording of the sample prompts that had been included in the trial version of the report. These reviewers responded to specific prompts as a whole, to portions of them, or even to single words in individual prompts on the basis of their own experience of having had students write to the same or similar prompts. Some described positive experiences; others reported problems. Some suggested changes in the wording of the topic or in the instructions. Others recommended abandonment of particular types. Representative comments follow.

Ruby Bernstein, a high school English teacher, pointed out that ETS argument topics or prompts don't elicit good writing, that students have a "livelier voice, have more to say when they are allowed to recreate experience." Jean Jensen reported that her ninth and tenth grade students responded very well to "story starters." Jensen also observed that certain prompts can be expected to elicit different levels of performance depending on grade level, ability level, and community (i.e. suburbia vs. inner city). Irving Peckham, a high school English teacher, also referred to the effect of age or stage of the student on response to certain prompts, but he observed that making up a story is an easier task for some individual students, regardless of age group or "stage of development," as posited in the study. Tanner commented that students tend to carry over their own patterns of speech and thinking into first person stories. He concluded that "they must be prompted to invent a character or a point of view different from their own natural propensities." Tanner said that teenagers will write stories, but "they tend always to use themselves as the central figure...if you give them a specific role which takes them out of themselves, they often flounder or write only briefly." Tanner also addressed the need for specification of plot:

Unless the prompt includes a clearly defined initial incident which has the potential, right in the prompt itself, of being a problem to resolve, one is unlikely to find many students, even students of creative writing, who can develop a plot. Students can develop a chain of happenings, but that is not necessarily a plot.

With reference to the content of the prompts, Kim Davis reported on certain prompts that Subject A has learned to avoid: education, writing, popular controversial issues, humor. On the first two of these, students write what they assume readers want to read; the third "produces a depressing blather of cliches;" and the fourth is "a subject so far removed from the circumstances of a writing exam that students are rarely amused by it." But the Subject A faculty do want students to write argumentative expositions, although they elicit it without resorting to "the dreary likes of 'abortion,' 'capital punishment,' 'marijuana,' etc." In the matter of political topics or prompts, Bill Robinson warned that the wording of a prompt should not force students to write from a bias they don't share. He advised caution, "remembering always that a school topic is not context free." Because students always know who will be reading their papers, Robinson said, using topics that elicit strong opinions could be "very touchy."

Tanner brought up an issue that no other reviewer considered: the role of knowledge in affecting the writer's point of view or stance, purpose, and vocabulary. As he put it, "writing is subject specific." A person must have something in mind to write about, some knowledge of a subject. And how that subject is developed affects the way the reader judges its merit. Tanner provided several illustrations of his point:

One can write about history only when he has learned some history and learned a bit about how history can be written. The subject makes demands on vocabulary. It makes demands on the writer's stance. It makes demands on what facts should be selected and how they should be put in order....Some people may want to use writing prompts to find out if a student can write about chemistry, or metal-working.

Tanner continued:

...when we holistic readers respond this way we are responding to subject matter sense rather than to mere writing skills. The irony is that one can write error-free prose about nonsense. When we read AP papers for ETS we are reading holistically for BOTH content and quality of expression. We are facing the SUBJECT SPECIFIC demands on any worthwhile writing....

....What I am saying is this: No 'transactional' prompt can be truly broadly successful without depending on reading skills or without being intentionally the culminating act to some body of study or learning experience. Hence in the nature of things one must evaluate such writing on a basis of subject matter as well as writing skills....From this monograph one would think that man can genuinely write in a vacuum of

reference, or at least that the reader can read a response without giving a hang whether or not it means anything.

Tanner also offered trenchant comments on the wording of prompts. He cited a case where he was question leader at the reading of an Advanced Placement English test (in about 1971 or 1972) which used a question based on a passage from an Orwell essay as a springboard:

The instruction went something like this: "Demonstrate how the author of this passage establishes his attitude toward the coming of spring." I was question leader on that question, and early on in the reading I was puzzled to have large numbers of papers brought to me with that question blank. Many obviously fine candidates, who did fine essays on the other two questions, simply skipped that first question entirely. And then I had a notion: I suspect that a great many students took that word "Demonstrate" to mean that they should pick subjects of their own and imitate Orwell's manner of expression. I was led to this belief by one paper in which a good candidate had actually done that as he "demonstrated" his own attitude toward the beach.

Tanner then gave an account of how he writes prompts to avoid such wording confusions:

I have since made it a rule of thumb for myself in attempting to write prompts that I would use an entirely limited vocabulary. I use "tell" when I mean either narration or description. "Tell what happened..." means narration; "Tell what the place looks like..." means description. I use "explain" when I mean either persuasion or exposition. "Explain how it works..." means exposition; "Explain what you believe and why..." looks toward persuasion. (Naturally I do not mean that I use those quotations; I mean merely that by using "tell" and "explain" and whatever appropriate phrases are needed after these commands, I establish the central mode or intention of the piece of called-for writing.)

Tanner illustrated how the substitution of "tell" for "describe" clarifies the meaning of a particular prompt, which could be interpreted to mean "describe an object which taught you a lesson." Tanner commented:

A student might scratch his brains for some time to invent a "lesson" taught him by some particular object-- a hard sidewalk or something--and end up with a rather silly piece of trumped-up writing. Try "Tell about an experience from which you learned a lesson," [or] "Tell about an activity you enjoyed doing as a child."

### Purposes of Writing Assessment and Rhetorical Purpose.

The issue of purpose was discussed in two contexts: Several teachers agreed on the importance of writing with a rhetorical purpose in mind--i.e., that the students should know what aim their writing is to accomplish. This notion also relates to the concept of audience. But several others--including some of the researchers and professional evaluators--addressed the concept of "purpose" in terms of the tester's reasons for having students write to these prompts, remarking that this objective hadn't been made clear in this preliminary version of the report. Tanner said:

Right now, I find only the vaguest things said about various purposes for testing of writing. If the paper is going to mention "purposes" shouldn't it show a list of purposes, the possible range of purposes? Shouldn't these then be discussed in relation to possible types of prompts which can be related to these purposes? I would try to discuss ways in which the wording of a prompt relates to any attempt at holistic reading.

Tanner seems to be asking about the function of the writing assessment presented in the trial version of the report. Is it intended to assess writing skills alone, or is it intended to assess writing skills as they can be brought to bear on the communication of some sort of subject-specific content?

With regard to purpose insofar as it includes audience, most teacher-critics agreed that it is useful to specify an audience; however, Peckham warned that students find certain instructions about audience to be "phony:" "This is the problem I've felt with the audience craze. Quite often a phony audience is as bad as no audience at all. Clear purpose seems more important. For expressive and poetic writing, no real audience is necessary."

And finally, with regard to the prompts, Kim Davis asked about the "reality" of students having to write to topics they'd just as soon not write to: "How should students be counseled to make the best of a bad situation?" After all, no matter how carefully the purpose for the writing and the audience for the writing are specified, students writing in a timed assessment situation are not going to produce "real" writing, no matter what the instructions read. What is being assessed, then, is not the students' ability to write in "real" situations, but the students' ability to perform in a highly artificially structured "test" situation, in which they agree to play the role of "real writers," and the teacher-reader agrees to play the role of a "real reader." In fact, they are really acting in their ordinary roles of student-performer and teacher-evaluator.

The Function of Instructions. Davis also agreed that the writing instructions are important--that it is the

instructions rather than the topic that determine the purpose of the writing. Peckham agreed that instructions for writing are best left simple. "Overwriting indicates uncertainty about the topic," he said. Jensen agreed fully with the need to avoid overprompting and underprompting in giving instructions. Tanner's anecdote about the misunderstanding of the term "demonstrate," mentioned above, illustrates the fact that sometimes a single word in the instructions can mislead the writer.

Two areas under the "Use and Misuse of Constraints" section of the trial version were discussed by several of the teacher-critics: Under the "Pick-A-Form Trap," Bernstein noted that the letter form has special problems. The "form" part takes up too much time and space, and, when used in her school, the results were boring. She called this an example of "testing for the wrong thing or putting the parts before the content." Robinson asserted that there were additional problems with asking students to pick a form, since this task of choosing a form added to the work the student was required to perform in a limited amount of time. Also, a given student may not pick the form that would have produced the best results for that particular topic. And most important for purposes of assessment, readers or evaluators have problems evaluating the same topic written in different forms. Robinson said, "I do not believe you can have a valid test situation in which students may pick different forms for their responses." (Several of the researchers and professional evaluators also questioned allowing the students to "pick a form" on a writing assessment task.)

Most of the teacher-critics agreed fully that specifying number of words, sentences, or paragraphs was not only an unhelpful constraint, but it adversely affected the writing that was produced. Peckham said of instructions specifying number of words: "They encourage a minimum response." And Bernstein pointed out that the ETS test booklet itself is a "constraint," since it provides the student with a limited and specified number of lines to fill up in 20 minutes.

Time Constraints. The fact that these prompts were being developed for use in timed assessments was questioned by three of the teacher-critics: Farrell stated that the study should not be limited to use with timed tests. Irving Peckham offered a special plea for enough time to rewrite: "We allow two periods at Live Oak. As well as stimulating rewriting in the classroom, our insistence on second drafts is a blessing for readers."

Bill Robinson also expressed concern about the effects of time pressure on the type of writing produced:

...the time allotted for response to a question has nothing to do with the mode of response to be elicited. That is, a speeded test will not promote expressive

writing, at least not in Britton's definition of expressive. What a speeded test will produce is writing more heavily characterized by oral-composing traits than a relatively non-speeded test. That is, the heavier the cognitive demands produced by the combination of response desired and time given, the more the response will resemble spontaneous speech, with all its "errors" and ungrammaticalities....the factor of time is one you haven't discussed as such, and yet it is, of course, very important.

Use of Pilot Tests. Two of the teacher-critics agreed with the need to pilot test the prompts before using them in a full assessment; however, both also stressed the importance of having teachers write to the prompts as well. Jean Jensen, a high school teacher, reported her experience at one school with pilot testing prompts with students: Despite their pilot testing, they still often received boring papers. At another school, where pilot testing was done among the teachers rather than among students, she reported that the student papers were better than at the other school, but she and her colleagues were not able to suggest reasons for this. Jensen stated, "I think it is tremendously important that these teacher/test-makers write on the prompt themselves and talk about the effect of a question on them....The teacher writing himself is crucial!...When teachers themselves also test the question--little trouble!"

#### Responses from the Panel of Researchers and Professional Evaluators

Ten researchers and professional evaluators submitted written reviews of Catharine Keech's trial version of "Practices in Designing Writing Test Prompts." These reviews consisted of 1- to 3-page letters giving general reactions to the whole report. Half of these respondents also provided from 10 to 30 textual annotations on the returned manuscript. As with the teacher-reviewers, most of the researcher/evaluator comments dealt with the content of the report, although some commented on the format, organization, and style. Though there are areas of common interest, the responses of the two review groups are not precisely parallel. The main themes and recommendations in these responses are collated below.

#### THE PANEL OF RESEARCHERS AND PROFESSIONAL EVALUATORS

1. Evans Alloway, Director of Test Development, Educational Testing Service
2. Gordon Brossell, Associate Professor of English, The Florida State University, Tallahassee
3. Rexford Brown, Editor of Publications and Test Developer, National Assessment of Educational Progress

4. Charles R. Cooper, Coordinator of Writing Programs, Department of Literature, University of California, San Diego
5. Paul Diederich, Senior Research Analyst, Emeritus, Educational Testing Service
6. Sarah W. Freedman, Assistant Professor of Education, Language and Literacy Division, School of Education, University of California, Berkeley
7. Ann Humes, Communication Skills Project Assistant, Southwest Regional Laboratory
8. Ina Mullis, Senior Research Analyst, National Assessment of Educational Progress
9. Joe M. Steele, Research Psychologist, College Outcome Measures Project, The American College Testing Program
10. Edward White, English Test Center, Department of English, California State University, San Bernardino

Value of the Study. The panel of researcher/evaluators generally praised the study as being greatly needed and for including many important and useful ideas. They considered it both helpful and interesting. Though several acknowledged the draft form, they also, nevertheless, pointed out the need for a careful revision, including a tightening of the style and organization. Charles Cooper, a writing researcher at the University of California, San Diego, supported the need for this study and applauded its achievement, but he considered the report to be too conjectural in some respects and indicated a need of more research to resolve some of the issues it raised.

Issues of Assessment Purpose. The interaction between assessment purpose and the nature of writing prompts to accomplish particular assessment aims was an issue with several of the respondents. Ina Mullis, Senior Research Analyst for the National Assessment of Educational Progress, said that clarification of the context, purpose, and audience for the report was needed. In reading the report, Mullis could not be sure what purpose the topics were intended to serve: instruction or assessment. If the purpose was assessment, then the purpose of the assessment was not clear. What kinds of decisions would a tester be making on the basis of the results? As indicated in the "Sources of Knowledge" section of this report, Mullis follows NAEP's "primary trait" system which calls for narrowly delineated writing exercises to assess particular accomplishments in writing according to rubrics established prior to testing. The approach in the Keech manuscript followed a much more open procedure amenable to holistic scoring and ranking of the quality of papers

according to rubrics developed during scoring.

Several of the researcher/professional evaluators, although supporting the study, raised questions about the purpose of the assessment procedure described in it. Gordon Brossell, a writing researcher at Florida State University, suggested that test makers need to determine what intellectual operations are to be assessed in addition to writing before they begin to construct a test. His implication--that it is not possible to test writing skills isolated from other skills--echoed Bernard Tanner's statements about subject-specific writing. Just as one of the pitfalls of writing prompts is to avoid making them into reading tests, so Brossell suggested that test makers determine what intellectual operations are to be assessed--in addition to writing--before they begin to construct a test. Sarah Freedman, a writing researcher, agreed that it was good to stress the importance of deciding what one wants to test before designing the task. Joe Steele, Research Psychologist for the American College Testing Program, asked for a clearer distinction to be made between writing instruction and writing assessment: "...evaluation of writing may call for a quite different approach and set of considerations than are most useful for instructional purposes." He said that he "found the attempt to deal with topics from elementary to college and for varied instructional and testing purposes to be too big a task to carry off well."

Brossell was unsure "why formal testing of writing in the elementary school needs to be done." If for diagnosis or evaluation, he asserted, these can be done without formal testing. He said this confusion may result from a blurring of the formal testing of writing with classroom writing assignments aimed at giving a teacher some evidence of a student's ability and progress in writing. It is interesting that while a number of the teacher-critics suggested that the focus of the study be broadened to include development of prompts for classroom instructional purposes as well as assessment purposes, several of the researchers and professional evaluators asked for a clearer distinction between writing prompts for the two purposes--of instruction and evaluation.

Need for Definitions. That there is not yet common agreement among researchers and professional evaluators on the usage of terms to describe the various aspects of writing assessment is evident in the responses of those who called for definitions. Gordon Brossell suggested a source of possible confusion--that research in the composing process of writers should not be applied to test situations unless such research involved the composing processes of writers in test situations. Joe Steele called the distinction between literature on scoring writing samples and directions for developing good topics artificial. He said the criterion for judging a topic "good" is the scoring procedure itself;

that the definition of a "good" test is (a) that it measures growth, and (b) that it "must discriminate, and do so in a range of performance that is of interest." Along these same lines, Paul Diederich, Senior Research Analyst Emeritus for the Educational Testing Service, gave the definition of a good topic, in which the criterion is not that it makes everybody write well; rather, "it is one that yields a wide standard deviation--from those who write with genius to those who can't write at all." Joe Steele suggested that one "criterion for judging that an essay test is good is that it measures writing that is valued as good outside the classroom." Two of the professional evaluator-critics referred specifically to aspects of holistic assessment that are different from other kinds of assessment. Rex Brown, Editor of Publications and Test Writer for the National Assessment of Educational Progress, stated that many of the problems and traps occur because the topics are being scored holistically rather than being keyed to specifications (e.g., primary trait guides), and the two kinds are not interchangeable. However, Ann Humes, Communication Skills Project Assistant at SWRL, stated that "one of the assets of holistic assessment is that prompts can be less explicit/complicated because...[one is] not assessing specific features."

And when it comes to defining and classifying types of writing prompts, as with the teacher-critics, several in this group criticized the use of the Britton classification system. Gordon Brossell said that becoming acquainted with a modern classification system probably is irrelevant to the actual task of writing prompts. Sarah Freedman said that the main problem was that Britton's categories don't fit. She said, "I'd like your topic classification scheme to capture different levels of cognitive demand that can be placed on writers by different topics."

The Content and Wording of Prompts. Several respondents criticized various aspects of the prompts, giving reasons and suggesting improvements. Ann Humes pointed out problems not only with the wording of some of the prompts, but also with the fact that it was not always clear what the intended grade level for each prompt was. In pointing out weaknesses in the wording of certain prompts, Edward White, English Test Center, California State University system, was critical of the inclusion of some of the examples for fear that they would become models of good practice for those who might not read the study carefully. Regarding types of prompts, Rex Brown suggested avoiding not only yes-no questions, but also "disingenuous pro and con questions." Ina Mullis agreed that story starters or endings should not be used with older writers, but she cautioned that "older" writers are not necessarily "more experienced" writers. Working with national samples, Mullis reported, NAEP found that 17 year-olds generally were not "experienced" writers. "They tend to be grateful for additional prompts, especially in a timed situation." (Note: Mullis probably means "instructions" instead of "prompts,"

in the context of this study's terminology.) She also cautioned against "opinion" topics with the interesting observation that writers who "agree" have a headstart on writers who "disagree," since the latter have to stop and take time to formulate their thoughts--to decide on what basis they disagree. Also, the more complex the initial statement, the more difficult the task will be for the writer who doesn't want to agree.

And finally, Rex Brown objected to the "trees" prompt: "It elicits too broad a range of responses. For me, the worst pool of papers for holistic scoring is a pool containing all discourse modes and many very different response strategies." He asserted that a topic for holistic scoring should minimize the number of dimensions along which the papers will be compared. (This would also seem to be an argument against allowing students to "pick a form.")

Purpose and Audience. Several of the respondents discussed the concepts of purpose and audience for writing. Ed White agreed with the study's emphasis on the need for specifying rhetorical purpose. Charles Cooper stated, "Specifying audience seems crucial to me in all explanatory and persuasion writing." Sarah Freedman asserted that "All test writing should by definition be audience-centered." However, Gordon Brossell said that providing "the full-rhetorical-context version of a test hindered rather than helped writers. Creating a naturalistic writing problem does not assure that writers will respond to it any better than to another version."

Functions of Instructions. With regard to instructions, Brossell noted that not only are their content and style important, but "the tone of instructions in tests can do something to alleviate test anxiety." Several other of the respondents discussed various aspects of the section on "Use and Misuse of Constraints." In the "Pick-a-Form Trap" section, echoing Rex Brown's earlier comment as well as responses from several of the teacher-critics, Sarah Freedman questioned giving students a choice of rhetorical form on the grounds that this would not result in a good ranked scoring. In the "Special Knowledge Trap" section, Freedman underscored the necessity of providing prompts based on universal experiences and allowing for choices within a given structure, as is the case with ETS, EPT topics, and junior proficiency topics at San Francisco State University.

With regard to specifying number of words, sentences, or paragraphs, Ann Humes, along with a number of others, agreed that it is best not to have such constraints, but she also noted that often a district will mandate that students be able to write a certain number of sentences, and this kind of policy, especially when it is part of a "skills curriculum," gives rise to this kind of specification. And although it is

generally agreed that skills such as sentence recognition, punctuation, and other mechanics can be measured using objective tests, Ina Mullis reported that "NAEP consultants have expressed the feeling that these skills should be measured by using actual writing samples."

Time Constraints. As with the teacher-critics, the question of timed tests was raised. Ed White mentioned the problem of "adequate time" for certain kinds of writing on timed tests, and said, "We still need discussion of appropriate time allotments for different tasks and different students." Brossell said that generally there is too little time provided, especially on college-level tests. But Brossell also went on to state that there needs to be more made of

the great difference between the writing situation in a classroom and that in a test. Good writing programs (and good writing classes) allow writers, even encourage them, to discover what they want to say and give them time and guidance in saying it. I submit that this is the antithesis of formal testing, where a writer is given one chance to do his best under constraints of both topic and time. Also, there's a difference between large-scale testing and evaluating writing performance in a classroom--especially from the writer's point of view....The psychological effects of the testing situation in my view are dramatically different from those of other writing occasions. The test is an isolated event in the lives of writers, often fraught with diabolical significance. Its immediacy and moment are quite imposing, even distracting. Just as a football player practices his plays diligently, then fumbles four times in the big game, so a writer can know all his moves but fail to execute under the pressure of a test. Tests need to be constructed with this kind of distinction in mind--the distinction between "practicing" and "playing"....Teaching writing and testing writing are much more distinct than they appear.

Use of Pilot Tests. Several of these respondents also mentioned the importance of pilot testing. Brossell stated that pilot testing is essential in the development of prompts--it provides empirical evidence for the effects of the prompts on writers. But he warned that pilot testing doesn't assure one that a prompt is doing what one wishes it to do. Interestingly enough, one of the professional evaluators, Ann Humes, not only agreed with the need to pilot test all prompts, but she also suggested pilot testing with teacher-colleagues first, and then with students.

#### Effects of Recommendations on the Final Version

Recommendations by Teacher Reviewers. Most of the teacher-reviewers made specific recommendations for changes

in the final version of the Keech manuscript on "Practices in Designing Writing Test Prompts." Many of the suggestions were incorporated into the final text. Some examples follow:

Mary Frances Claggett, a high school teacher, suggested that there be more examples of under- and over-prompting and other forms of unhelpful constraints which practitioners could use as a reference in their own development of prompts. Such examples were included in the final version.

Bill Robinson noted that writing about an imaginary situation "can produce the Great Conditional Problem," and suggested that this idea be expanded. The final version discusses the problem of the conditional tense with regard to imaginary situations, and proposes specific wordings of this kind of prompt in order to avoid eliciting the conditional.

Tanner suggested providing more examples of responses to sample prompts, and this, too, was done to a limited extent in the final version of the report. However, samples of responses to prompts are to be found in later chapters of the full report. See particularly, the chapters relating to the studies done at Los Gatos High School and for Santa Clara County.

Tanner also suggested providing an ill-conceived prompt and then showing its revision, with reasons for the revision and explanations for one working and one not working. Along these same lines, Bernstein suggested that including sample rubrics would be useful. Although she agreed that it is important to discuss papers at length before scoring them, she said it was also important to have written criteria for determining scores. In this context, she made an interesting observation about different rubrics being needed for competency tests, proficiency tests, and advanced placement tests, because of the different natures of each kind of test; that is, each of these represents a different kind of testing intent. The Tanner suggestion was acted upon. Poorly conceived prompts are introduced with recommended revisions and discussions of the problems. Ruby Bernstein's recommendation seemed to go beyond the scope of this phase of the study, and thus it was not acted on.

An interesting suggestion came from four of the eight teacher-critics who responded. Farrell suggested "an heuristic in the form of a set of questions that any teacher or test maker might ask himself/herself about a writing prompt." Such questions, Farrell noted, could also be an efficient way of summarizing the major points made in the study. Tanner also saw the need for an heuristic:

What are the questions that a writer of a prompt should ask himself as he tries to write a prompt for a certain type of test? What are the issues? How do these issues

relate to the needs of the reader who will be trying to read the written responses and sort out the levels of success these who show toward the target objectives?... I would try to list these questions and concerns. I would try to relate these concerns to the language of the prompt, and to the purposes of the test.

On the practical level, Bernstein suggested an anecdotal example of a school's or a department's step-by-step procedure for developing a prompt--how the teachers went about the formulation of it, from start to finish, including pitfalls and problems. Claggett underscored the need for such an example with her own recounting of how her district writing center has used "brainstorming" to come up with prompts. Their trial-and-error approach to using them has had only a 50-50 success rate. Some sort of heuristic would be better than this "haphazard approach," she said.

At least three sets of guidelines have been included in this final report. In an appendix to Part II there is the set of Guidelines for writing prompts devised by Gertrude Conlan, Chief Reader for the Educational Testing Service. In Part I, there is a synthesis of principles for constructing prompts derived from the recommendations of teachers and scholars. There is also in Part I a set of guidelines for writing prompts based on a recorded session where three teachers met to develop an assessment prompt for an actual assessment.

Robinson suggested that because many teachers and test administrators might not be familiar with "testing talk," there should be included in the study some brief definitions of certain terms, especially "valid" and "reliable," with emphasis, as he put it, on "the inherent unreliability of essay tests" to measure writing ability. Along these same lines, he said that a section was needed on "ways in which the reliability of the essay test can be increased--or the unreliability compensated for."

Keech has included in her final version of Part II a definition of key terms, and throughout Parts I and II there are discussions of the limitations of various forms of measurement of writing competence.

A final suggestion for improving the design of writing exercises came from Jean Jensen, who suggested adoption of the rule followed by the teachers in her school: "If the papers you get are a disappointment, then what you wanted was not what you asked for. Discuss the 'topic' with the class." One of the main studies in the Writing Assessment Project was devoted to investigating the match between the task expectations of the tester and the student writer's interpretation of the task provided. The results of this investigation are presented in Technical Report No. 4.

Recommendations by Researchers and Professional Evaluators. Fewer specific additions were suggested by the researchers and professional evaluators; for the most part, they responded to what was there. Charles Cooper suggested that more illustrations be provided of specific points, and this was done in the final version. Rex Brown suggested that "it might be useful to explain some of the problems and traps in terms of the end of the process--holistic scoring--rather than the beginning. Perhaps you could present a description of the ideal pool of papers for an ideal holistic scoring by ideal scorers, and then--having established the principles of the end process--work backwards toward the principles of topic creation." Somewhat along these same lines--and similar to the suggestion made by several of the teacher-critics--Ed White stated that "the most important lack...is a description of the process of question development." Ann Humes likewise recommended: "Use a heuristic device in writing a prompt to make sure you have covered the relevant rhetorical elements; e.g., A. what is the purpose (why?), B. what is the subject/topic (what?), C. who is the audience (who?), D. what is the form/genre (how?)."

## MORE RESULTS OF THE STUDY

The following set of guides for designing writing test prompts are presented here with the possibility that they may aid their users in developing understanding of the nature and range of effect that a prompt may evoke. We need to do our best to create prompts that at least touch the edge of the student writer's world knowledge. We need to offer them starting points for composing that enable them to move freely within whatever kinds of guiding frameworks are given. We must be sure that the prompt is couched to forestall the needless difficulties that arise from ambiguous language and confusing signals. We may never write the perfect prompt, but pupils deserve to meet year-by-year assessments prompts that are as exactly principled as possible in their design.

### Guides for Designing Writing Test Prompts

The following guides are presented in the state of development they had reached by the end of the Writing Assessment Project. They await further collaborative refinement as fellow professionals test them in practice. The set below includes:

1. The Philosophy of the Assignment for Writing Assessment
2. Guidelines for Developing Topics for Writing Assessment
3. A Tentative Schema of Writing Tasks for the Development of Assessment Prompts
4. Schema for Task Analysis of Prompts

#### THE PHILOSOPHY OF THE ASSIGNMENT FOR WRITING ASSESSMENT\*

**Maxim:** A well-framed assignment for a writing assessment provides an occasion for writing in which the student's powers of expression and communication are stimulated to their normal maximum.

**Principles:**

1. The subject chosen must be potentially interesting to the teacher-reader or evaluator of the essays written.
2. The subject chosen must be potentially interesting to the student writers.
3. The assignment furnishes data to start from which is open to presentation in any of several forms ranging from statements or questions to pictures or cartoons accompanied by instructions.
4. The assignment must be meaningful within the student's experience.
5. The assignment seeks to elicit a specific response and limits content or form or both.
6. The assignment is itself specific in suggesting an audience beyond the teacher or evaluator when the subject makes such specification feasible.
7. Assignment by topic alone is artificial and yields a lack of focus in development.

**Coda:** A well-framed assignment, clear and precise in its instructions, shows awareness of the audience of waiting examinees. The demands on the assignment creator are identical with those on the writer.

Leo Ruth

*\*A synthesis of professional wisdom based on my report on SOURCES OF KNOWLEDGE FROM THEORY AND RESEARCH FOR DESIGNING WRITING TEST PROMPTS. NIE/BAWP Writing Assessment Project, University of California, Berkeley 1982*

## GUIDELINES FOR DEVELOPING TOPICS FOR WRITING ASSESSMENTS\*

(These guidelines are based on a topic development session with Ruby Bernstein, Claire Pelton, and Bill Thomas participating. BAWP Writing Assessment Project Study of Topic Development. University of California, Berkeley, May 1981.)

### Some Questions to Consider in Creating Topics for Writing Assessments

1. Why are we doing this? What is the purpose of this writing assessment?
2. What will we do with the results?
3. What is the nature of the student population to be assessed?
  - a. What is their grade level?
  - b. What is their general level of academic competence?
  - c. What prior curricular experience have they had that prepares them for this assessment?
4. What is the most appropriate topic subject for this assessment?
  - a. What prior knowledge will discussion of this subject require?
  - b. Is this subject accessible to a wide range of students?
  - c. What is the relation of the topic subject to the students' world?
    - Is it connected to real student concerns?
    - Is it remote or alien to their world?
    - Is it an interesting or involving topic?
  - d. Are there aspects of the subject that may lead to stereotyped thinking, reliance on clichés?
  - e. Is there any element of the subject that may be considered unduly threatening to personal values, religious beliefs, self-esteem, etc.?
  - f. Is the subject free from biasing effects for different sub-groups within the student population?
    - Is it too limiting for upper level ability students?
    - Does it presume experience that may have been denied some students for social, cultural, or economic reasons?
    - How difficult might the task be for the lowest twenty-five percent of the student population?
  - g. Will the topic lead to diatribes against persons at the school, parents, local citizens?
  - h. How is the subject related to the school curriculum and can equal exposure to similar curriculum experiences be assumed?
  - i. How does the choice of subject affect choice of rhetorical mode?
5. How should the topic subject be worded?
  - a. What do the verbs used communicate to the student? e.g., Describe? Tell? Explain? Illustrate? Analyze? Discuss?
  - b. What do the commonly used formal labels actually communicate? Essay? Story? Argument? Persuasive essay? Letter? Journal entry?
  - c. Should a range of formal options be proposed?
  - d. To what extent should students be cued? e.g. "Write about one thing..." "Tell about three improvements..."
  - e. What level of background detail should be provided?
  - f. Should the subject include specification of an audience?
6. How should instructions be handled? Embedded in the topic or kept separate? Should the topic contain hints about approaches to the subject or suggestions for task management?

\*Prepared by Leo Ruth

## A TENTATIVE SCHEMA OF WRITING TASKS FOR THE DEVELOPMENT OF ASSESSMENT PROMPTS\*

### GENERAL WRITING PURPOSES

1. To describe
2. To narrate
3. To record/report
4. To explain/direct
5. To persuade
6. To request
7. To plan
8. To edit

### WRITTEN PRODUCTS TO BE ELICITED

- 1.1 Sustained description of person, place, or object
- 1.2 Description and expression of feelings toward what is described
- 1.3 Sequenced description from memory
- 1.4 Description to accompany given pictures
- 2.1 Imaginative narrative based on given characters and setting
- 2.2 Original end to a story selected by the pupil
- 2.3 Short story
- 3.1 Autobiographical account of event experienced recently
- 3.2 Verifiable account of an event
- 3.3 An account of something learned
- 4.1 Explanation and reflection on a regulation
- 4.2 Explanation of a complex skill
- 4.3 Instructions to accomplish a simple household task
- 5.1 Informal letter to a friend to get reader to change mind
- 5.2 Argument justifying a point of view
- 6.1 Letter to a person in a public/private agency
- 6.2 Letter of application
- 7.1 Account of an activity to be undertaken
- 8.1 Editing of a written account

Note: All tasks envision provisions for variation in audience and the writer's control over appropriate forms. The framework provides for attention to four dimensions:

1. Variations in mode: narrative/descriptive; reflective/analytical
2. Degree of control granted writer or assessor: ranging from open to closed specifications
3. Source of subject matter: first-hand experience or knowledge from elsewhere
4. Range of purposes: functional and literary

It is expected that analytic marking following upon an initial holistic scoring would take into account the variations in content, structure, style, and grammar that different writing tasks impose. It is also assumed that topics would be conceived under the general philosophy accompanying this schema.

\*Adapted from Gorman, T.P., et al. Language Performance in Schools: Primary Survey Report No. 1 (1981), p. 85, and Secondary Survey Report No. 1 (1982), pp. 66-67. London: Her Majesty's Stationery Office. These two reports present the results of the 1979 assessment conducted by the Language Monitoring Team at the National Foundation for Educational Research in England and Wales. Leo Ruth, NIE/BAWP Writing Assessment Project, University of California, Berkeley, 1982.

SCHEMA FOR TASK ANALYSIS OF PROMPTS\*

LEO RUTH

PROBLEM: WHAT RANGE OF INFORMATION IS REQUIRED IN A WRITING TASK STATEMENT TO ENABLE A WRITER TO PROCEED WITH PRECISION—TO RECOGNIZE AND ACCOMPLISH THE TASK SPECIFIED IN A MANNER EXPECTED? TASK SPECIFICATION IS VIEWED AS A PROCESS OF THE REDUCTION OF UNCERTAINTY.

CATEGORY            1)PRIOR KNOWLEDGE                            2)GENRE                            3)NOTION  
PARADIGM DRAWING ON ANY EXPERIENCE OR IMAGINATION, WRITE SOMETHING ABOUT SOMEONE OR SOMETHING  
1. \_\_\_\_\_  
2. \_\_\_\_\_

CATEGORY            4)AUDIENCE                            5)ROLE                            6)SYNTAX                            7)FUNCTION  
PARADIGM TO SOMEONE, EMPLOYING ANY VOICE OR REGISTER IN ANY STYLE FOR ANY PURPOSE OR USE  
1. \_\_\_\_\_  
2. \_\_\_\_\_

CATEGORY            8)QUANTITY                            9)RESULT  
PARADIGM AT ANY LENGTH TO ACHIEVE ANY EFFECT.  
1. \_\_\_\_\_  
2. \_\_\_\_\_

TOPIC/TASK STATEMENT: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

\*NIE/Bay Area Writing Project Writing Assessment Project, University of California, Berkeley

## REFERENCES

- Abelson, Robert P. Constraint, construal, and cognitive science. Proceedings of the third annual conference of the cognitive science society. Berkeley: University of California, 1981.
- Applebee, Arthur N. A study of writing in the secondary school. Urbana, Illinois: National Council of Teachers of English, 1980.
- Atwater, James D. Better testing, better writing. Papers on Research about learning. New York: Ford Foundation, 1980.
- Ballard, P.B. Teaching and testing English. London: University of London Press, 1939.
- \*Bartlett, E., 1981
- Bassey, Michael. Pedagogic research: on the relative merits of search for generalisation and study of single events. Oxford Review of Education, 1981, 7, No. 1, 73-94.
- Bereiter, C., Scardamalia, M., and Bracewell, R.J. Writing and decentered thought: the development of audience awareness. Paper delivered at American Educational Research Association, Toronto, Canada, March, 1978.
- Braddock, Richard, et al. Research in written composition. Champaign, Illinois: National Council of Teachers of English, 1980.
- Braddock, Richard. Evaluation of writing tests. In Reviews of selected published tests in English. Urbana, Illinois: NCTE, 1978.
- Britton, J. N. Multiple marking of english compositions: an account of an experiment. London: The Schools Council, Her Majesty's Stationery Office, 1966.
- Britton, James, et al. The development of writing abilities (11-18). London: Macmillan Education, 1975.
- Brossel, Gordon. Rhetorical specification in essay examination topics: an experimental study. Unpublished paper. Tallahassee, FL: Florida State University, 1981.

\*See Addendum

- Brown, Rexford. Choosing or creating an appropriate writing test.  
In L. Kasdan, D. Hoerber (Eds.) Basic writing. Urbana: NCTE, 1978.
- Cameron, William Bruce. Informal sociology. New York: Random House, 1963.
- Capell, Frank J., and Quellmalz, Edys S. Empirical validation studies of alternative response modes for writing assessment. (CSE Report No. 145). Los Angeles: Graduate School of Education, University of California, Los Angeles, 1980.
- \*Carr, A., 1965
- Clark, C.M. and Florio, S. Understanding writing in school: a descriptive study of writing and its instruction in two classrooms. (No. 104) Lansing, MICH: Michigan State University, 1982.
- Clark, Michael. There is no such thing as good writing (so what are we looking for?) in A. Freedman and I. Pringle (Eds.) Reinventing the rhetorical tradition. Conway, ARK: L & S Books, University of Central Arkansas, 1980.
- Coffman, William E. Essay examinations. In Educational Measurement. Washington: American Council on Education, 1971.
- Commission on English. Freedom and discipline in English. New York: College Entrance Examination Board, 1965.
- Commission on Writing. A quest for question. In English Series. New York: The Educational Council, AEVAC, Inc., 1966.
- Cooper, Charles. Writing abilities of regularly-admitted freshmen at SUNY/Buffalo. New York: University Learning Center and Department of English, SUNY/Buffalo, 1979.
- Crowhurst, M. and Piché, G.L. Audience and mode of discourse effects on syntactic complexity in writing on two grade levels. In Research in the Teaching of English, 1979,13,101-109.
- Crowhurst, M.C. Syntactic complexity in two modes of discourse at grades 6, 10, and 12. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1978. (ERIC Document Reproduction Service No. ED 168 037).

\*See Addendum

- D'Angelo, F. J. Modes of discourse. In G. Tate (Ed.) Teaching composition: 10 bibliographical essays. Fort Worth, Texas: Christian University Press, 1976.
- Elvin, Lionel. The place of educational research. Oxford Review of Education, 1975, 1, NO.3, 189-204.
- Farr (Whiteman), Marcia. In A. Humes (Ed.) Moving between practice and research in writing: Proceedings of the NIE-FIPSE Grantee workshop. Los Alamitos, CA: SWRL, 1981.
- Farrell, Edmund J. The beginning begets: making composition assignments. English Journal, 1969, 58, No. 3, 428-431.
- Fillmore, Charles J. Ideal readers and real readers. Paper delivered at Georgetown University Round Table, Washington, D.C., 1981.
- \*Flower and Hayes, 1980  
 Flower, Linda S. Problem solving strategies for writing. New York: Harcourt, Brace and Jovanovich, 1981.
- Flower, Linda S. and Hayes, John R. A process model of composition. Washington, D.C.: Document Design Project (Tech. Report No. 1) National Institute of Education, 1979.
- Foley, Joseph J. Evaluation of learning in writing. In Handbook on formative and summative evaluation of student learning. Benjamin S. Bloom (Ed.). New York: McGraw-Hill, 1971.
- Freedman, Aviva and Pringle, Ian. Why students can't write arguments. Ottawa, Ontario, Canada: Linguistics Department, Carleton University, 1981.
- Godshalk, F.I., Swineford, F. and Coffman, W.E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.
- Gordon, Raymond L. Interviewing: strategy, techniques, and tactics. Homewood, Illinois: The Dorsey Press, 1980.
- Graves, Donald H. Balancing the basics: let them write. Papers on research about learning series. New York: Ford Foundation, 1978.
- Graves, Donald H. Research doesn't have to be boring. Language Arts, January, 1979, 56, 76-80.

\*See Addendum

- Greenberg, Karen L. Some relationships between writing tasks and students' writing performance. Paper delivered at the 71st Annual Convention of the National Council of Teachers of English, Boston, 1981.
- Greenberg, Karen L. The effects of variations in essay questions on the writing performance of CUNY freshmen. New York: Instructional Resource Center, City University of New York, 1981.
- Guy, William. Pedagogic Explanation. Cambridge Journal of Education, 1975, 5, No. 1, 21-29.
- Harpin, William. The second 'r': writing development in the junior school. London: George Allen and Unwin Ltd., 1976.
- Hoetker, James. Effects of essay topics on student writing: a review of the literature. Unpublished paper. Tallahassee, FL: Florida State University, 1981.
- Hoetker, J.; Brossell, G.; and Ash, B. Creating essay examination topics. Unpublished paper. Tallahassee, FL: Florida State University, 1981.
- Hogan, Padraig. A critique of educational reserach. Oxford Review of Education, 1980, 6, No. 2, 141-155.
- Irmscher, William F. Teaching expository writing. New York: Holt, Rinehart and Winston, 1979.
- Humes, Ann. Specifications for composition instruction. (Technical Note No. TN 2-80/21), Los Alamitos, California: SWRL Educational Research and Development, 1980.
- Kandell, I.L. Examinations and their substitutes in the United States. New York: The Carnegie Foundation for the Advancement of Teaching, 1936.
- Kennedy, Graeme. The language of tests for young children. In Bernard Spolsky (Ed.) The language education of minority children. Rowley, Massachusetts: Newbury House Publishers, Inc., 1972.

- Kincaid, Gerald L. Some factors affecting variations in the quality of students' writing. Unpublished Ed.D dissertation, Michigan State University, 1953.
- Kinneavy, James L. A theory of discourse. New York: W.W. Norton and Company, 1971.
- Kinneavy, James L. A pluralistic synthesis of four contemporary models for teaching composition. In A. Freedman and I. Pringle (Eds.) Reinventing the rhetorical tradition. Conway Arkansas: L & S Books, University of Central Arkansas, 1980.
- Kirrie, Marjorie. Prompt writing is not impromptu. NWP Network Newsletter, May 1979, 1, No. 3, 6-7
- Labov, William. The study of language in its social context. In Joshua A. Fishman (Ed.) Advances in sociolinguistics of language. 1971, 1, 152-216. The Hague, Holland: Mouton.
- Lanham, Richard A. Revising prose. New York: Scribner's Sons, 1979.
- Lloyd-Jones, Richard. Primary trait scoring. In C. Cooper and L. Odell (Eds.) Evaluating writing: describing, measuring judging. Urbana, IL: National Council of Teachers of English, 1977.
- Long, Russell C. Writer-audience relationships: analysis or invention? College Composition and Communication, May, 1980, 31, 221-226.
- Meier, Deborah. Reading failure and the tests. New York: The Workshop Center for Open Education, February, 1973.
- Michigan Survey Research Center. Interview Manual. Michigan, 1969.
- Miles, Josephine. Working out ideas: predication and other uses of language. (Bay Area Writing Project Curriculum Publication No. 5) Berkeley, CA: University of California, 1979.
- Miller, Susan. Rhetorical maturity: definition and development. In A. Freedman and I. Pringle (Eds.) Reinventing the rhetorical tradition. Conway, ARK: L & S Books, University of Central Arkansas, 1980.

- Moffett, J. and Wagner, B.J. Student centered language arts and reading, k-13. Boston: Houghton Mifflin Company, 1976.
- Moffett, J. Teaching the universe of discourse. Boston: Houghton Mifflin Company, 1968.
- Moffett, J. Active voice. Montclair, New Jersey: Boynton/Cook, 1981.
- Mullis, Ina V.S. The primary trait system for scoring writing tests. Washington, D.C.: U.S. Department of HEW, National Institute of Education, 1976. (ERIC Document Reproduction Service No. ED 124 942).
- Neilson, Brooke. Writing as a second language: psycholinguistic processes in composition. Unpublished dissertation. San Diego, University of California, 1979.
- Nold, E. and Freedman, S. An analysis of readers' responses to essays. Research in the Teaching of English, 1977, 11, No. 2, 165-174.
- Odell, Lee. Teachers of composition and needed research in discourse theory. College Composition and Communication, 1979, 30, 39-44.
- Odell, Lee, Cooper, Charles R. and Courts, Cynthia. Discourse theory: implications for research in composing. In Research on composing: points of departure. Cooper and Odell (Eds.), Urbana, Illinois: NCTE, 1978.
- Oppenheim, A.N. Questionnaire design and attitude measurement. New York: Basic Books, Inc., 1966.
- Park, Douglas B. 'The meanings of "audience"'. College English, 1982, 44, No. 3, 247-257.
- Payne, David A. and McMorris, Robert F. Educational and psychological measurement. Waltham, Massachusetts: Blaisdell Publishing Co., 1967.
- Payne, Stanley L. The art of asking questions. Princeton, New Jersey: Princeton University Press, 1951.

- Paulin, Pauline. Organizing for holistic scoring. California English, 1980, 16, No. 1, 26-27.
- Pearce, John. Examinations in English language. Schools Council Programme in Linguistics and English Teaching Papers Series II, 1974, 4, 35-81. London: Longman Group Ltd.
- Perron, J.D. The impact of mode on written syntactic complexity: part III - fifth grade. Studies in Language Education (Report No. 27), University of Georgia: Department of Language Education.
- Pianko, Sharon. A description of the composing processes of college freshmen writers. Research in the Teaching of English, 1979, 13, No.1, 5-24.
- Rockas, Leo, Modes of rhetoric. New York: St. Martin's Press, 1964.
- Rosen, H. An investigation of the effects of differentiated writing assignments on the performance in English composition of a selected group of 15/16 year old pupils. Unpublished Ph.D dissertation, University of London, 1969.
- Rowntree, Derek. Assessing students: how shall we know them? London: Harper and Row, 1977.
- Ruth, Leo P. Questions and answers: a model of relations, forms, and functions. Unpublished dissertation. Berkeley, CA: University of California, 1980.
- Sax, Gilbert. Principles of educational measurement and evaluation. Belmont, California: Wadsworth Publishing Co., 1974.
- San Jose, C.P.M. Grammatical structures in four modes of writing at fourth-grade level. Unpublished dissertation. Syracuse, NY: Syracuse University, 1972.
- Sanders, S. and Littlefield, J. Perhaps essays can reflect significant improvement in freshmen composition. Research in the Teaching of English, 1975, 145-153.
- Scott, F.N. The teacher and his training. In Carpenter, Baker and Scott (Eds.) Teaching of English. New York: Longmans, Green, 1903.

- Shuy, Roger. Closing remarks. In A. Humes (Ed.) Moving between practice and research in writing: proceedings of the NIE-FIPSE grantee workshop. Los Alamitos, CA: SWRL Educational Research and Development, 1981.
- Smith, Laura Spooner. Measures of high school students' expository writing: direct and indirect strategies. (CSE Report No. 133), Los Angeles: UCLA Graduate School of Education, 1980.
- Stahl, Abraham W. Structural analysis of children's compositions. Research in the Teaching of English, 1974, 8, No. 2, 184-204.
- Starch, D. and Elliott, E. The reliability of grading work in English mathematics, and history. In D. Payne and R. McMorris (Eds.) Educational and psycholological measurement. Waltham, MASS: Blaisdell Publishing Co., 1967.
- Staton, Jana. Introduction. In E. Bartlett, Learning to write: some cognitive and linguistic components (Linguistics and Literacy Series 2). Washington, D.C.: Center for Applied Linguistics, 1981.
- Steinberg, Erwin R. A garden of opportunities and a thicket of dangers. In Gregg and Steinberg (Eds.) Cognitive processes in writing. Hillsdale, New Jersey: LEA, Inc. 1980.
- Stenhouse, Lawrence. What counts as research? British Journal of Educational Studies, 1981, 29, No. 2, 103-114.
- Volpe, Richard. Knowledge from theory and practice. Oxford Review of Education, 1981, 7, No. 1, 41-51.
- Whale, K.B., and Robinson, S. Modes of students' writing: a descriptive study. Research in the Teaching of English, 1978, 21, 349-355.
- Wilkinson, Andrew. Assessing language development: the Credition project. Washington, D.C.: U.S. Department of HEW, National Institute of Education, 1980. (ERIC Document Reproduction Service No. ED 178 906).

Winters, Lynn. The effects of differing response criteria on the assessment of writing competence. (CSE Report No. 131), Los Angeles: UCLA Graduate School of Education, 1980.

Wood, Dorothy Adkins. Test construction. Columbus, Ohio: Charles E. Merrill Books, Inc. 1960.

#### ADDENDUM

Bartlett, Elsa Jaffe. Learning to write: some cognitive and linguistic components (Linguistics and Literacy Series 2). Washington D.C.: Center for Applied Linguistics, 1981.

Carr, Arthur J. A student writing assignment based on "Fire Walking in Ceylon." In Commission on English kinescripts (Set I). New York: College Entrance Examination Board, 1965.

Flower, Linda and Hayes, John R. The dynamics of composing: making plans and juggling constraints. In L. Gregg and E. Steinberg Cognitive processes in writing. Hillsdale, N.J.: Lawrence Erlbaum Ass., 1980.

Wilkinson, Andrew, et al. Assessing language development. Oxford: Oxford University Press, 1980.

APPENDIX

A

List of Invited Critic Readers  
for Technical Report #1, Part II.

Evans Alloway	Director of Test Development, Educational Testing Services, Princeton, NJ
Kate Blickhan	Language Arts Coordinator, Tamalpais Unified School District, CA
Gordon Brossell	Associate Professor of English, Florida State University, FL
Rex Brown	Editor and Writing Test Developer, National Assessment of Educational Progress, Denver, CO
Gertrude Conlan	Chief Reader for College Boards, Educational Testing Services, Princeton, NJ
Charles R. Cooper	Coordinator of Writing Programs, Department of Literature, University of California, San Diego, CA
Paul Diederich	Senior Research Analyst, Emeritus, Educational Testing Services, Princeton, NJ
Edmund J. Farrell	Department of Curriculum and Instruction, College of Education, University of Texas
Linda S. Flower	Carnegie-Mellon University, Pittsburgh, PA
Sarah W. Freedman	Language & Literacy Division, Graduate School of Education, University of California, Berkeley, CA
Ann Humes	Communication Skills Project Assistant, SWRL, CA
Marjorie Kirrie	Department of English, Portland State University, OR
Ina Mullis	Senior Research Analyst, National Assessment of Educational Progress, Denver, CO
Edys Quellmalz	UCLA Center for Study Evaluation, Los Angeles, CA
Joe M. Steele	Research Psychologist, College Outcome Measures Project, American College Testing Program, Iowa City, IA
Edward White	English Test Center, Department of English California State College, San Bernardino, CA

APPENDIX

B

Bay Area Writing Project  
Teacher-Consultants

Ruby Bernstein	Northgate High School Concord, CA
Owen Boyle	Language & Literacy Division, University of California, Berkeley, CA
Mary Frances Claggett	Alameda High School Alameda, CA
Kim Davis	Director, Subject A University of California, Berkeley, CA
Dorothy Facchino-Letcher	American High School Fremont, CA
Mildred Gamble	St. Vrain Valley Public Schools Longmont, CO
Kent Gill	Davis High School Davis, CA
JoAnne Graham	San Mateo Elementary District San Mateo, CA
Jean Jensen	Las Lomas High School Walnut Creek, CA
Nancy McHugh	Grant High School Sherman Oaks, CA
Irving Peckman	Live Oak High School Morgan Hill, CA
Bill Robinson	Dept. of Fresman English California State University, San Francisco, CA
Bernard Tanner	Palo Alto High School Palo Alto, CA
Ken Williams	Modesto City Schools Modesto, CA
Dan Wolter	Vintage High School Napa, CA

# APPENDIX

C

UNIVERSITY OF CALIFORNIA

BERKELEY · DAVIS · IRVINE · LOS ANGELES · RIVERSIDE · SAN DIEGO · SAN FRANCISCO



SANTA BARBARA · SANTA CRUZ

**BAY AREA WRITING PROJECT**

Supported by

National Endowment for the Humanities  
Carnegie Corporation of New York  
California State Department of Education  
University of California, Berkeley  
Bay Area Schools

BERKELEY, CALIFORNIA 94720

January 29, 1981

Bernard Tanner  
2591 Ramona  
Palo Alto, CA 94301

Dear Bernard:

As part of our Bay Area Writing Project study of Topic Effects in Holistic Assessment, we are now preparing a Technical Report on topic development from the point of view of teachers, researchers, and professional evaluators. The report will accumulate illustrative "good" and "poor" topics along with sage advice on the problems in topic development and how to avoid them. Our Research Coordinator, Catharine Keech, has prepared a preliminary version of such a report based upon her experience as a member of the Bay Area Writing Project Evaluation Team and upon anecdotal material supplied by other BAWP Staff and Teacher Consultants.

We are sending you a draft of this report on topics, hoping that you will agree to serve as a Critic-Reader of the paper. From the perspective of your long experience in preparing topics for writing assessments, we hope you will comment on your perceptions of its completeness, accuracy, organization, and style of presentation. We are particularly interested in receiving anecdotes from your own experience relating to the unexpected behavior of certain topics. We also invite you to offer observations about what makes "good" or "poor" topics. We expect to incorporate into the final version of the report a summary of the comments that we receive.

We hope you can quickly review the paper we are submitting. To save time, we suggest you simply annotate and return particular pages or sections of the paper that you wish to comment on in the envelope provided. If you are unable to undertake this assignment, please return the manuscript in this envelope.

We can pay a modest honorarium of fifty dollars (\$50.00) for this review, which will be sent as soon as we receive your comments with your social security number.

Sincerely yours,

Leo Ruth, Director  
NIE/BAWP Writing Assessment Project

149

131

PRACTICES IN DESIGNING WRITING TEST PROMPTS:  
ANALYSIS AND RECOMMENDATIONS\*

Catharine Keech

INTRODUCTION

The recent effort to introduce direct measures of writing performance into local writing assessments has found local teachers and administrators with few resources to guide them in developing good essay tests. While there is an extensive literature describing procedures for creating valid and reliable objective tests of certain kinds of learner achievement, there is very little in the way of science to guide the development of performance tests, especially tests of writing ability. This is not to say we do not have a growing literature on how to score writing samples, with advocates of several competing methods addressing themselves to some extent to the problem of developing good prompts or writing stimuli with instructions for the writer. In fact, perhaps the most extensive work taking into account the relation of the test-topic to the scoring plan has been done by developers of what is called Primary Trait Scoring, (Nullis, 1974; Lloyd-Jones, 1977). But as with other approaches, the title of this approach to assessment identifies it as mainly a scoring approach. ETS test developers, like all evaluators attempting to develop good direct measures of writing ability, are willing to admit for the most part that writing good topics remains more an art than a science. Up to now, deciding when a topic is good has been typically more intuitive than objective.

This report raises a question basic to the conduct of writing assessment: How should students be asked to write and what should they be asked to write about if we wish to measure their real writing abilities? The assigned writing task plays a critical role in determining how students perform and whether testers are measuring what they think they are measuring.

Creating a good prompt is not easy. Students write best when they write with intention. That is, they must be able to find something they want to say to someone. However, "intention" is not the same as "motivation." No matter how highly motivated they may

---

This report is based upon research supported by the National Institute of Education under Grant No. NIE G-50-0034 to the Bay Area Writing Project, University of California, Berkeley, Leo Ruth, Project Director.

be to perform well on a writing test, no matter how concerned they are to achieve a good score, they will not be able to do their best if they find the topic dull, confusing, or intimidating. They need to seize on the germ of an idea and begin writing with confidence if they are to generate a complete piece of writing in a limited amount of time. The intent to communicate releases abilities which motivation leaves untapped.\*

In creating a writing prompt, the tester's main purpose is not to single out clever students who can decipher complicated instructions, but rather to make it easy for as many students as possible to respond enthusiastically, demonstrating their best writing abilities. The test-maker cannot always be sure that students will interpret the task as intended and respond appropriately: sometimes a single word will mislead a writer--sometimes instructions are inadequate to inspire a sustained piece of writing. Testers need to be sure that every student will be able to respond, but they are also interested in whether the kind of writing elicited will allow measurement of the abilities they wish to evaluate. Different kinds of prompts elicit very different kinds of writing--some students do better at one kind than another; some teachers and administrators are more interested in one kind than another.

In addition, while a classroom teacher can discuss the assignment with students, check their understanding of the task, and provide additional instructions as needed, the test-maker must write the assignment in a form that is as nearly fool-proof as possible. The assignment in a test often must guide the writing of students from many different classrooms, including students accustomed to different interpretations of such commands as "describe," or "write a short essay about...." The test-maker has recourse to several strategies to offset these difficulties, the most important of which is pilot-testing the writing assignment or checking on its workability with a sample group. (Kirrie, 1980.)

The following review of practices in designing prompts aims to identify basic principles that guide professional test-makers, to alert novice test-makers to potential problems in prompt design, and to explore the relationship between the nature of the writing test

---

\*See Woodward and Keech, 1980 (pp. 6-10) for an extended discussion of the distinction between motivation and intention.

and the purposes of the assessment. The first section clarifies terms used in the discussion of writing assignments and distinguishes between two basic parts of a prompt: 1) the topic or other writing stimulus and, 2) the instructions to the writer.

The second section focuses on the nature of the topic or other stimulus. It relates the choice of topic to the test purposes and the kinds of writing desired, as well as pointing out several common problems for topic writers.

The third section deals with the instructions to the writer and how these relate to the criteria by which the writing is to be judged, as well as to the test-maker's decision about what is and is not being tested. It relates the wording of these instructions to the purposes of the assessment, and it points out some common problems in giving instructions to writers. A list of suggestions for writing essay prompts and for judging the worth of a writing prompt as prepared by Gertrude Conlan of the Educational Testing Service is included in Appendix A. A list of the writing tasks discussed in this report with their sources appears as Appendix B. References for this report follow the appendices.

## THE NATURE OF THE WRITING PROMPT

### A Note on Terms

Multiple choice tests of verbal skills are often labeled writing tests or even composition tests although no writing or composing is involved. In this report, however, the term, writing test, will refer only to those tests which call for one or more samples of student writing. This general term, writing test, is preferred to the traditional use of the expression essay test, which is reserved here for those writing tests which call on students to produce essays, per se. The term, writing sample, refers to the writing produced for writing tests, whether by a single student or by an entire student population. (In this report, there is an alternation of reference to student-writers and writers though the full term is always intended.) Hence, the designing and administering of the writing test is followed by the collecting and scoring of the writing sample. (Generally, those who score the writing samples are simply called readers, but in order to avoid confusions with other common usages of this term, the more cumbersome form teacher-readers is used frequently in this report.) These personnel and procedures together constitute the holistic writing assessment.

The invitation to write in a writing test has been referred to variously as the topic, the task, the prompt, the test question, the writing stimulus, the writing instructions, the assignment. In this report, some of these terms will be used interchangeably-- especially the writing task, the prompt, the test question-- to refer to the whole writing assignment. That assignment generally consists of two main parts: 1) the topic, subject, or stimulus--the tester's identification of what the writer will write about; and 2) the instructions--the particular suggestions about or limitations on the content of the writing, and the tester's constraints on how the writer is to address the subject, with any specifications on what the writer is expected to do.

### The Structure of the Prompt: Topic and Instructions

In designing the writing prompt, the test-maker must choose or invent topics or other stimuli that provide students with something to write about; formulate instructions for writing which help students perform the writing task to the best of their ability; and ensure that the topic and the instructions work together to elicit the particular writing skills the tester is interested in measuring. We shall consider first the selection of topics and then the role of

instructions with other types of stimuli.

In a writing test, the student's composition is expected to address the given topic, and it will be judged in part on how well it addresses that topic. The student's response to the topic will be matched against the interpretation of the topic expected by the evaluator. All topics can be considered prompts for writing, but not all writing prompts provide a fully formed topic for the writer to develop into an essay.

When the prompt is just a word or a phrase, it requires the student to generate a predication which forms the topic into a subject for development.

Josephine Miles (1977) has noted that the giving of a topic, in the sense of the giving of a subject to write about-- for example, "My Home Town"-- does not provide the writer with a thesis or a generalization which might generate a coherent essay. The writer must discover his own thesis by shaping a predication for his subject that is susceptible to demonstration and development; for example, "My home town provided a rich environment for the development of a bad character."

The conception of "topic" which requires students to generate their own predications is exemplified in A Thousand Topics for Composition: Revised, collected in the Illinois English Bulletin (1971), or in the more recently published What Can I Write About? Seven Thousand Topics for High School Students (Powell, 1981). Most of these topics are short phrases which might serve as titles of student essays: "my home town," "a childhood memory," "my favorite object," "sex education," "capital punishment," "making tacos," "the case for/against dieting."

The nature of some writing tests, however, makes it more appropriate to speak of a "stimulus" rather than a topic for writing. The stimulus for writing could be a famous quotation, a complete poem, a controversial statement, a fragment of an essay or story, a photograph or drawing, a piece of music, a taped conversation, a film-clip, a news bulletin. In these cases, the instructions to the writer rather than the stimulus alone, will determine how the prompt serves to elicit writing. The student writer might be directed to respond to a photograph, for instance, by creating an imaginative piece which recounts a personal experience, relates a stream of consciousness, or tells an invented story with no reference whatever to the photograph itself. But when the student is asked to describe, comment on, criticize, evaluate, agree or disagree with, explain, or in some way to write about

the given stimulus, then a line of development has been indicated and the stimulus plus instructions become the full prompt for writing.

For younger children or for tests of general fluency, very simple phrasings of both topic and instructions may be quite effective in eliciting intentional and coherent writing. For example:

PROMPT 1a

Think about the word trees. Write about what the word suggests to you.\*

Note that the topic points to one word: "trees." In this case, the instructions to the writer are quite simple and provide almost no constraints: "Think about the word..."; "Write about what the word suggests...." The same topic could be accompanied by more elaborate instructions which would encourage students to move out of writing only for self-expression toward writing for literary, or even for academic or social purposes:

PROMPT 1b

Think about the word trees. Write a story in which trees play an important part. This may be a real experience or an imaginary story. You may even write as if you were a tree. Be sure that trees play some part in your story.

PROMPT 1c

Think about the word trees. Write an essay in which you explain one of several possible ideas this word may suggest to you. You may write about a special kind of tree, a special use of trees, people's attitudes toward trees, or something else that interests you about trees. Focus on one idea--do not attempt to list all the things trees can mean.

In the above two examples, it is not the topic but the writing instructions that determine whether the test is one of general fluency calling for self-expression and free exploration of internal feelings.

\*All credits for prompts appear in the Prompt Index, Appendix B.

or whether it is a test of more refined composing strategies in which specific forms or discourse conventions will be rewarded. In example 1a, instructions are brief and very simple; in examples 1b and 1c, instructions have been more fully specified.

With a single-word prompt, the instructions chosen by a tester can be repeated from year to year, test to test, without changing the demands of the task, simply by replacing the stimulus word, with equally familiar terms for the population tested, for example, masks, scars, plastic, sports, work, play, fire. The more open to individual interpretation, the more likely the topic is to elicit intentional writing from students. The wide range of responses elicited by such open topics also makes interesting reading, combating the reader fatigue that sometimes results when hundreds of predictably similar essays must be scored. The next section discusses the selection and wording of the writing topic to specify the content of the essay. (Methods of selecting and presenting quotations, literary passages or non-verbal stimuli such as pictures or music are not discussed.) The final section addresses the problem of writing good instructions for different test purposes.

## THE WRITING TOPIC: CONTENT AND FORMS

### Section Overview

Testers select and phrase the writing topic with three different concerns in mind: the need to guarantee accessible content, the need to specify or limit content, and the need to elicit particular kinds of writing. The topics discussed below will be evaluated for their ability to guarantee every student something to say, for the manner in which they specify what students write about, and for their appropriateness in eliciting different kinds of writing or writing for different purposes. In the following discussion, rough distinctions are made between four kinds of writing tasks.

First, introspective or autobiographical writing is done for purposes of exploring and sharing the writer's personal world, and is useful for measures of general fluency. The second kind of writing task is discussed below under "Topics That Draw Upon Personal Experience." Although testers are advised against trying to distinguish between fact and fiction in student narrative, they may examine a student's narrative abilities either by appealing to the students' own experiences or by appealing directly to their imagination, asking them to create an imaginary world, with topics calling for or allowing fictional or fantasy writing. This kind of task is discussed under the heading, "Topics That Elicit Made-Up Stories." Some of the topics listed as appealing to personal experience tend to elicit primarily descriptive as opposed to narrative writing. Other topics which attempt to elicit purely descriptive writing, without reference to a more general rhetorical aim, are discussed in "The Writing Instructions: Purposes and Problems," in the final section of this report, under the sub-heading "The Describe Trap."

Third, one might offer topics which call for business writing or social writing, both done to accomplish transactions typical of the world outside of school. Business writing topics such as complaint letters or job applications are discussed also in the final section of this report under the sub-heading, "The Special Conventions Trap." Social or community-directed writing includes arguments or persuasive writing about public issues. Topics for assessing argument skills or persuasive writing are examined below under the sub-headings, "The Problems of Prior Knowledge Requirements in Exposition and Argument Topics" and "Controversial Issues as Topics." Although writing is also done in school settings which is aimed at persuading others to agree with one's values, to adopt certain attitudes, or to take certain actions, all such writing is considered

social in function, of interest to testers when both writing and reasoning skills are to be measured.

And finally, we come to expository writing, which is typical of writing done in an academic setting, and is intended to explore ideas and concepts about a world which is shared by reader and writer, to interpret events, or to make explicit the meaning of experiences. Expository writing is also discussed in conjunction with social or persuasive writing, but it differs from the latter in that the writers are not asked to write as if they wish to affect a reader's actual behavior in the real world or change the course of events, but rather as if they wish to project an idea for examination and evaluation by the reader as part of a shared process of making sense of experience. These rather abstract distinctions between argument and exposition may be ignored for most testing purposes, since the thinking and language skills needed for these two kinds of task are often similar.

It should be pointed out that not only the topic but also the instructions to the writer will determine which of these kinds of writing is likely to be elicited. Introspective or autobiographical writing, in particular, may range from almost pure narration of events to almost pure commentary, the content of exposition. The framing of the topic and instructions can encourage the writers in either direction--toward a concrete representation of an experience, or toward a reflective and explicit examination of the ideas derived from an experience; or they can leave the writers the full range of possibilities in their own construction of the task, as when the test is one of general fluency rather than of specific narrative or expository skills. The remaining parts of this section deal in turn with specifying content for personal experience writing, exposition, and argument.

### Topics That Draw Upon Personal Experience

Popular topics that provide the writer with rich content and elicit relatively fluent, honest writing close to the student's natural voice are those which encourage introspective or autobiographical writing by drawing upon personal experience. Such writing is easy to motivate and is highly intentional. When someone asks us what we would wish for most in the world, or what we recall that was happy or sad or surprising from our childhood, or whom we especially admire, we believe the questioner to be genuinely curious. We recognize this rhetorical situation from oral speech, and most of us welcome an invitation to explore our own minds or re-examine our experiences in conversation with a

friendly listener. Unlike topics which require students to write about information which they believe the reader already has, these topics allow students to impart information which only they possess: on what they remember, what they think, how they feel they can write with authority.

Topics designed to elicit personal expression rarely specify the precise subject of the discourse; rather, they specify the predication.

PROMPT 2a

NOT

Write about December 25, 1973.

PROMPT 2b

BUT

Write about an experience you had during a family holiday which was particularly memorable. It may be happy or sad, but should be a time when you had strong feelings.

Instead of naming a specific event or character, as one might in a test about literature, the tester invites the writer to select material which fits certain specifications, the most important usually being that the event or character is memorable in some way. The underlying rhetorical task is to present the experience or portray the character in a way which demonstrates that this person, object, or event is actually worth writing about. Although the writer might comment, explain, or analyze, such commentary is not essential if the writer is sufficiently skilled in descriptive and narrative techniques to make the experience memorable for the reader.

The following topics have been used to elicit interesting and fluent introspective or autobiographical writing:

PROMPT 3

Write about a place that is special to you. Help your reader understand its peculiar interest or significance. The place you choose may be of any kind--for example, a personal haunt, a historical spot, or an institution, or a place you remember vividly

from early childhood.

PROMPT 4

Write about an object you are especially attached to, something that has deep personal meaning for you, something that has become a part of your life. You might want to consider the way you discovered it, the way it came into your life, the way it has taken on meaning through time.

PROMPT 5

Tell about something you enjoyed doing as a child.

PROMPT 6

Tell about an experience from which you learned a lesson.

PROMPT 7

Think of someone you admire or resent. (Not your parents.) Write about this person and relate one or two incidents that led you to admire or resent this person, so the reader can understand why you feel the way you do.

PROMPT 8

Describe a change that has happened in you recently or an event that happened a long time ago but still affects you. Tell of the events that surround it and caused the change in you so your reader can understand what is now different about you and why.

Students may respond to these topics using varying proportions of description, narrative, or commentary. Numbers 3 and 4 are most likely to evoke descriptive writing, while number 5 is more likely to inspire a narrative. Although numbers 5 through 8 may elicit entirely narrative accounts, they may also evoke discussion and commentary of the kind that develops from expository topics, as the writer examines his subject from a distance and attempts to draw conclusions or

reflect on the meaning of his experience. The researcher's experience suggests that all but the more advanced or mature writers will use narratives or anecdotes whenever they are allowed to do so. To be sure of eliciting commentary and explicit discussion of ideas, the tester will need topics that are more like those used by the College Board English Composition Test (See also pp. 56), such as:

#### PROMPT 9

Each of us assumes a personality to satisfy each group into which we enter; i.e., we do not act at home precisely as we act with the gang, in the classroom, or before an employer.

This latter prompt models the generalizations that form the basis for expository writing, but it encourages students to draw on their own concrete experiences in exploring the truth of the predications advanced in relation to the topic. The addition of instructions to the writers helps translate the quotation into issues the writers might recognize in their own lives.

Teachers find that prompts which allow students to write directly from their own feelings and experiences are excellent for assessments of general writing abilities, because such prompts point to sources of meaningful subject matter for most writers. The wording of the prompt can be changed for different age groups without changing the underlying rhetorical problem. Prompts 3 and 4 above, for instance, could read simply:

#### PROMPT 10

Write about your favorite place.

#### PROMPT 11

Write about your favorite object.

Prompts which demand that students shift from writing directly about their own feelings and experiences to writing about ideas and information have so far been used primarily with older students. Current work in cognitive psychology suggests that students below the twelve to fourteen year old age range may have limited ability to sustain abstract discourse involving formal definitions, reasoning, analyzing. On

the other hand, young children can often write with great involvement about concepts and information encountered in history and science activities, as well as writing about their own values or the nature of God. Using autobiographical topics is the best way to test for fluency, self-expression, and mastery of the written code, since personal topics provide the best guarantee that the specified content will be equally accessible and interesting to all writers.

Cuing Personal Topics for Content. Choosing interesting topics and expressing them simply will ensure that most students can find something meaningful to say in their writing. But some of the above topics include content specifications which "cue" students to see the full potential of the topic, helping them recognize relevant material from their own lives, and clarifying the tester's expectations. In the following autobiographical prompt, the cuing is underlined:

PROMPT 12a

Think of an occasion when you experienced something for the first time. It may be something you later came to take for granted, but for some reason, the first time was memorable. Write about your experience so that your reader understands your feelings at the time and why it stays in your memory.

This prompt has elicited interesting writing from high school students ranging from pure narrative to informal exposition.\*

In another assessment, the prompt was more elaborately cued. (Woodworth and Keech, 1980.) Testers may create however many cues they think will be relevant to their students' experiences:

PROMPT 12b

Think of an occasion when you experienced something for the first time.

CUES:

There is a book called The First Time, in which famous people describe their first kiss or first love affair. But there are many other important First Times in our lives,

---

\* See Chapter 2.

from the first time we tied our own shoes, or went to school or camp, to our first time in deep water, or first big game or first speech in front of a group. Some of these occasions are first and only times. Others are the first of many times, but for some reason the first time was memorable.

**INSTRUCTIONS:**

Choose one of these occasions that stands out in memory and write about it so that your reader can understand your experience and why it was memorable.

**PROMPT 12c**

Think of an occasion when you experienced something for the **FIRST TIME**. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing.

The process of cuing substitutes for class discussions which enable students to think of experiences in their own lives they might wish to write about. The cuing may clarify the kind of experience the tester expects the student to write about, explain the writer's options, or simply stimulate ideas as in the following prompts:

**PROMPT 13**

Compare two stores you know which sell similar items. Consider such things as prices, merchandise, service, atmosphere, location, or other special features. Which of the two stores do you prefer? Explain why.

**PROMPT 14**

Compare two actual holidays or special events you celebrate in your family (such as

Thanksgiving, Passover, Christmas, the Fourth of July, your birthday, etc.) Consider such things as food, decorations, activities, behaviors, atmosphere, or other special features. Which of the two do you prefer? Explain why.

Elaborate cuing has two dangers, however. First, poor readers may be discouraged by the sheer amount of text in the writing prompt, even to the point of not writing, or writing off-topic. Second, unless the cuing is carefully phrased, it may seem to the student to offer commands rather than suggestions. The following prompt, recommended by a state agency for use in elementary schools (California State Board of Education Handbook on Writing Assessment, 1979) includes highly directive cuing, intended to ensure that students write at some length and in comparable ways, but it imposes strong limits on what students can choose to write about:

#### PROMPT 15a

You have met a man from outer space who has landed on earth near your school playground. He can understand English, but he does not know anything about school here on earth. Describe your school for him. Tell him about your school building. Tell how the school building looks on the outside and the inside. Tell him about your teacher and your classmates. (italics added)

The prompt has several unfortunate characteristics. Using a space man as audience serves no purpose that would not be better served by a human audience, someone who had a believable reason to be interested in a written description of the school. More serious problems, however, occur in the cuing for content, which gives students too much to do, preventing them from focusing and writing with more depth about one or two aspects of the school. (See discussion below on "Focusing".) The cuing emphasizes the physical building, which is perhaps the least interesting and most difficult thing to write about. (See discussion below of the "Describe Trap".) Finally, the imperative tone of the prompt allows the students no freedom to choose which parts of the school are significant to them and which they would like to write about. The top-scoring student papers that were offered as examples were extremely limited in what they attempted and generally lacked focus and coherence. The following alternative

is suggested:

PROMPT 15b

Imagine a friend of yours is about to move to town. Your friend's name is Pat. Pat will be going to your school. Below, write a letter to Pat, describing one or two things you think a new student should know. You might tell about what you do at lunch time, or describe the games you play. You might describe your favorite part of the classroom or what your teacher allows you to do in class activities.

This version has not been pilot-tested, so it cannot be asserted that it is more successful than the former version in eliciting interesting writing. On the other hand, this wording represents a more natural rhetorical framing of the task, putting the student in the role of intentional communicator rather than that of direction-follower. Because the framing of a test question often becomes a model for the teacher's framing of a classroom writing assignment, the test prompt should represent, as closely as possible, the kinds of rhetorical situations in which students are encouraged to shape their own discourse meaningfully. Where students have had little practice in writing, a tester may find they write longer papers in response to prompts such as 15a above which lists a set of things to mention. But the more naturalistic framing of Prompt 15b could well provide students with genuine opportunities to write.

The following prompt, from a California elementary school assessment, raises useful guiding questions about what students should include and how their essays might be shaped, but it had an unfortunate side effect on some inexperienced writers:

PROMPT 16a

Write about your favorite place. Name the place and tell where it is. Tell how often you go there. Why is this place special to you? How do you feel when you are there? What are some sights, sounds, colors and smells in your favorite place?

All the questions in this prompt provide excellent ways to stimulate ideas during brainstorming or prewriting sessions. In fact, the cues in this prompt were used

for a prewriting exercise on a sheet of paper which listed each command or question followed by blank spaces for the student's response. During the prewriting, and again on the test day, some students perceived the writing test as a short-answer quiz and wrote words and phrases rather than sentences in response to the commands or questions. To many students, the following type of response seems acceptable, given the phrasing of the prompt:

#### WRITING SAMPLE

Disneyland. Hollywood. Once a year.  
Because it is fun. I feel good. Happy.  
Tired. Sights: Mickey Mouse, rides, African house. Sounds: loud roars, lots of people.  
Colors: red, yellow, blue, all colors.  
Smells: cotton candy, caramel apples, animal smells, people smells.

It is important then to make the prompt sound less like a set of questions to answer and more like an invitation for communication, as in this revision:

Everyone has a special place where they like to go. Write about your favorite place to go. Describe the place so your reader can imagine what it is really like. Tell about what you do there, and how you feel when you are there. Help your reader see why this place is special.

In summary, when testers desire to give students assistance in finding and selecting personally meaningful content, they should consider methods of cuing which avoid sounding too much like directions to the writer and more like genuine options. Cues should be arranged if possible to suggest a natural way of planning the response to the topic. Elaborate cuing is appropriate only if students are competent readers and have adequate time to consider the suggestions before beginning to write. Elaborate cuing removes some of the work of composing for poor writers but may hinder good writers. In fact, the best writers will often regard the listed options as things not to write about, thinking that "everyone else will write about those," and seeking instead to find a subject more novel than those suggested by the tester. For tests of general writing ability which use personal experience topics, a well-phrased topic and simple cuing are usually adequate to stimulate content in the resultant writing.

### Topics That Elicit Made-Up Stories

Literary writing is the writing of fiction, autobiography, and poetry. Some children will begin to use literary techniques quite early in responding to the topics listed under personal writing, above. On the other hand, if it is story-telling abilities per se that are being tested for, prompts that draw on writers' direct experience would have to be replaced by prompts that challenge their imaginations, calling for fantasy or fiction stories.

Frequently stories are elicited by prompts which give starter lines or conclusions. A middle school included the following task as one part of its writing assessment:

#### PROMPT 17

Write a story that begins: 'One day I found a magic machine.'

Other successful starters have been:

#### PROMPT 18

This morning on my way to school, I saw a big black hole.

#### PROMPT 19

The last thing I remember was the ship in flames, people screaming and running for lifeboats. Then I blacked out. When I came to, I was...

The instruction part of these prompts should specify that the examiner is calling for the creation of a narrative so that the student's sense of story will give structure to the piece. In these pieces, flashes of personal insight or memories or real feelings that arouse the reader's sympathy will be less important than story structure--imaginative incidents, vivid details, conflict, pacing, coherence, good closure. Story-writing is a special ability and hence requires a specialized test. There is no reason not to score such a test holistically, but one cannot use such test topics as if they were parallel in task demands to the autobiographical topics. The topics offered to inspire writing about personal experience, while they often elicit literary strategies, may also encourage

Pretend I am the Wish Fairy who can make your wishes come true. I want you to write me a letter to tell me what you wish for. Think of several things you wish for. (pause) Now then, choose the one thing you wish for the very most.

Tell me what it is by finishing the sentence, "I wish I had..." What do you wish you had? After you write down its name, then go on writing and tell me all about it. What is it like? Write down everything you think and feel about your wish.\*

The instructions above can easily be modified for parallel versions of the "wish" prompt:

#### PROMPT 20b

I wish I were...  
I wish I could...  
I wish I knew...

In the early years of holistic assessment at a Bay Area high school, the teacher test-makers were primarily concerned about creating prompts that students would enjoy writing about in the annual assessment. At that time they were not using the annual assessment for placement purposes or even program evaluation. The following parallel prompts, used in their first three annual testings, were felt to be highly successful in eliciting writing that was both competent and interesting:

#### PROMPT 21

Write about an event you wish you had witnessed or could witness. The event can be real or imagined; the time of the event can be past, present, or future. Make it clear why the event is significant to you. (1973)

---

(\*The advantage of oral instructions in elementary or even junior high is that guidance and prewriting stimulation is provided without transforming the writing test into a reading comprehension test. In holistic assessments administered in more than one class, however, teachers should read a standardized set of oral instructions to avoid variations in the task description.)

use of language patterns that are similar to those of oral language. Hence the student may attend more to the problems of transcribing these ideas "aloud" in writing, than to coping with the additional constraints of trying to create a made-up story.

Making up a story is possibly an easier task for children at some stages of development than is writing about their feelings, wishes, or experiences. For younger writers, writing means writing stories. They may not make up very good stories, or recognize when they have a beginning, middle, or end, but they sometimes write a great deal more in the attempt than if the tester had asked them, for instance, to describe a favorite object. It may be that they understand the rhetorical intent behind the command, "write a story," whereas they do not recognize the rhetorical intent behind the command, "tell me about someone you admire." On the other hand, one elementary school teacher points out that students need far more time and space to develop a story. They often write many pages which they are reluctant to re-read or revise. (Gentry, 1980.)

Starter lines or conclusions are not effective for use with older students. The more experienced writers often rebel at intervention in their creative processes. Requests to write out of personal experience or to write about imaginary situations are more likely to be honored as legitimate prompts. A good example of such a topic is the job application which is discussed below, in the last section under "The Special Conventions Trap".

#### Hypothetical Situations as Subjects for Writing

Many teachers find that topics which ask students to imagine, pretend, or wish often inspire lively and interesting writing. An effective elementary school test consists of the following words on a piece of blank paper:

#### PROMPT 20a

Dear Wish Fairy,  
I wish I had . . .

The teacher gives the following instructions orally:

PROMPT 22

If you had to choose to be something other than a human being, what plant or animal or other form would you choose? In your writing, give your reader some idea of what you think it would be like to be that form, and of why you chose it. (1974)

PROMPT 23

If you could change places with someone else, who would it be? The person you write about can be living, dead, drawn from past or present, from books, films, etc., or from your own imagination. In your writing give your reader some idea of what it would be like to be that person, and of why that life appeals to you. (1975)

However, these three topics require writers to project themselves into hypothetical situations. A major problem with these hypothetical topics is that this led many students to begin their responses using conditional verb forms, such as "I would...", "I could..." Only the very best high school writers who began in this way were able to maintain a consistent point of view throughout their essays. Robinson (1981) and Kirrie (1980) have both warned against the use of hypothetical situations for testing writing skills because of the need for students to use conditional or subjunctive verb phrases. As Robinson says, "When students are put into situations where they have to use the conditional, they invariably demonstrate that they cannot."

However, each of three prompts listed above (21, 22, 23) ended with these instructions:

You may do this writing as a journal entry, character sketch, dialogue, letter, story, autobiographical essay, argument, poem, or other form.

Even though many testers advise against offering students such a wide range of choices (see the discussion of the "Pick-A-Form Trap" that follows), these alternative modes often provide students with a way of avoiding what Robinson calls the "Great Conditional Problem" when they are confronted with hypothetical situations.

Many students in the assessments where the hypothetical topics were used avoided the conditional altogether by plunging directly into narratives or by using purely descriptive writing. For the prompt that asked them to imagine being non-human, some students adopted the voice and persona of the non-human being and wrote a kind of interior monologue, e.g., a beetle reflecting on "my life as a beetle."

The test's purpose must determine whether to use imaginary situations as writing prompts. The experts cited advise against using imaginary situations as prompts; however, if one wishes to test the student's ability to sustain the conditional point of view, prompts such as the following are useful:

#### PROMPT 24

What would you do if you had a million dollars?

The following prompts describe imaginary situations which proved interesting to student writers, in a project designed to pilot test a number of topics.\* They elicited mostly narrative and descriptive writing free of the conditional problem because these prompts placed students inside the situation rather than asking them to comment on what would, should, or might happen might happen if such a situation occurred:

#### PROMPT 25

While visiting in your grandmother's house you find a large trunk. Propped on a chair beside the trunk is a card; on it in your grandmother's handwriting are the words: "To my only grandchild: Inside this trunk you will find what you have always wanted." Describe the gift and how you will use it.

#### PROMPT 26

Imagine that you are trapped in one of the following situations and you cannot escape for several hours. Tell what you do to escape or to pass the time until you are rescued:

\*Santa Clara County Writing Assessment Field Test of Prompts in Chapter 3.

- A. an elevator
- B. an abandoned mine
- C. a locked room
- D. a life-boat

PROMPT 27

Imagine that you have just landed on the moon. Write about what happens after you climb out of your space ship and set foot on the surface of the moon.

The conditional problem does not seem to hamper younger writers who usually have strong feelings about fairness and a good sense of what they might do in an imaginary situation. Thus they often respond enthusiastically to questions that begin, "What should you do when...." The following exercise has been used with fifth and sixth graders as an informal classroom activity.

PROMPT 28

Pretend you write the advice column in our school newspaper. You must answer the following letter:

Dear Happy Helper:

I have a little sister who always hits me when I have a toy or a cookie she wants. My grandmother says I can't hit her back. She says I should share, but I get tired of sharing. What do you think I should do?

Signed,

In Trouble

The research finds that responses to this prompt will generally consist of conditional forms such as, "You should....," "You might...." or imperatives, such as, "Do this." Thus, students benefit from prewriting discussions during which they can practice giving reasons for their advice. While other prompts are better tests of narrative abilities, this one tests the ability to make suggestions or give advice.

Imaginary situations that present moral dilemmas ("What would you do if...") can become loaded questions if the teacher readers actually believe that there is really only one right answer. For instance, in the above prompt, if readers accepted only non-violent solutions to the problem, students who made other suggestions would be scored down. If students were not told in advance that there was only one correct kind of response, this would become a "loaded question."

In scoring writing from prompts such as the above which involve moral dilemmas or value choices, teacher readers naturally tend to judge some answers as more "right" or more admirable than others. However, in holistic assessments of general writing abilities, evaluators probably desire to judge whether the students can communicate clearly what they think a person should do in a given situation, not whether their choice seems "good" or "bad." Thus trainers generally try to help readers resist letting moral judgments strongly influence scores--unless such judgments are included in the goals of the assessment.

#### Problems of Prior Knowledge Requirements in Exposition and Argument Topics

For exposition and argument the question of particular prior knowledge becomes more involved in the writing task. The better informed writer, the writer who has thought about an issue, is likely to produce more cogent arguments, more convincing examples to support his or her position or to explain his or her thesis. Some writing tasks have failed to elicit good writing even from good students when none of the students tested had adequate information about the topic presented for discussion, or when too many of the students had not previously considered a problem in sufficient depth to write any but the most simplistic and unreflective comments. The following topic was eliminated from one assessment after teachers tested it by attempting to write on it themselves:

#### PROMPT 29

Explain your agreement or disagreement with the following statement: The history of women in our country is the history of oppression.

Two teachers admitted that they did not know enough about the history of women to write a sound essay on the subject. The topic would be appropriate, of course, if all students being tested had just completed a course on the history of women in their country. In

that case the topic would test course content mastery in addition to writing ability; teacher readers still would need special training to reach agreement on what constituted an adequate response to the topic.

In one large assessment that was scheduled to be held over two years beginning during the drought in California, testers pilot-tested a topic that asked students to describe the effects that drought had on the campaign urging people to conserve water, and to make specific recommendations about how to do so. Student responses were well-informed and imaginative, often passionate, but the topic had to be dropped because testers could not be sure that the second year of the assessment would provide a drought or a parallel crisis as certain to have touched the lives of all students in the sample.\*

To some extent the expository or argument topic which requires students to marshal evidence from the real or literary world is a test of general knowledge. It becomes a test of special knowledge only if content is narrowly specified, as it must often be to shape a persuasive topic. A test of general writing ability need not discriminate in favor of students whose special interests or background give them access to information not widely shared, if the topic allows students to argue directly from their individual experiences and/or reading. Better informed students will have an advantage in any case, of course, but the choice of topic can minimize or maximize that advantage, depending on the purpose of the assessment: whether the writing task is intended primarily as a skills test or as a special content test.

To avoid problems of special knowledge requirements, testers can provide a body of relevant facts or a literary passage to be commented on or argued from. Writing from given information may test reading, analysis, and reasoning, as well as writing abilities, but no student is put at a disadvantage for not having encountered specific facts or read specific literature.\* (See Prompt 31a) Another option is to provide a

---

(\*Note: Carnegie funded evaluation of Bay Area Writing Project, Michael Scriven, Director, 1979.)

\*This kind of prompt is used in the military and business worlds on writing tests for job advancement, and resembles performance measures called "In-basket tests", where candidates are asked to respond to the kind of problems likely to appear in their "In-baskets." This is also a common format for college placement tests like the Subject A exam at University of California, Berkeley which provides an essay or

stimulus--quote, word, picture, idea, etc.--and ask students to draw on their own knowledge or experience to shape their essays. Testers must give clear instructions to students to draw on their reading and experience for examples to explain their theses. This is the common format on many college placement exams. (See pp. Prompts 34 and 35 below.) Again, test-makers may provide a rhetorical situation sufficiently common that every student can imagine facing such a task and allow students a range of alternatives, including imaginary ones, in their solution to the problem. A good example of such a topic is the job application situation presented in Prompt 43.

### Controversial Issues As Topics For Persuasive Writing

Controversial issues, such as abortion or capital punishment, seem to provide especially appropriate topics for tests of persuasive writing. Students can be expected to have much to say about either side of a current controversy; however, their papers may reflect unconsidered parroting of opinions heard in the media or from parents. Controversial topics can create special problems for holistic assessors if the subject arouses strong biases in both student-writers and teacher-readers and distracts readers from making a careful evaluation of the student's writing abilities.

Topics based on issues of sexism, for instance, often inspire heated and honest writing, but they also may lead to expressions of hostile, sexist views on the part of both males and females. Such views can antagonize teacher-readers who feel equally strongly about the issues. Special training can overcome these responses to some extent, but for most test purposes, it may be wise to avoid direct statements of extreme positions, such as the following:

#### PROMPT 30a

Men are superior to women.

#### PROMPT 30b

Women are superior to men.

---

literary passage for critical analysis, thus testing reading and analytic skills to be taught in Freshman English courses.

INSTRUCTIONS:

Agree or disagree with the statement, giving reasons for your opinion.

A different, slightly more sophisticated, phrasing appeals less to prejudice and more to observations and experiences, thereby inspiring more interesting and scorable writing (Caplan and Keech, 1980):

PROMPT 30c

Women are better off than men in today's society.

INSTRUCTIONS:

Agree or disagree with the statement, giving reasons for your opinion.

An entirely different approach to this topic was taken by an evaluator who wished to test the students' understanding of logic and their ability to recognize good and poor arguments (Keech, 1978):

PROMPT 30d

INSTRUCTIONS:

Explain what is wrong with the following argument: "It's pretty clear that on the whole, women are superior to men. In the physical dimension, the best evidence for this is simply that women live longer, and by a substantial number of years. Other evidence is that women are virtually immune to several diseases that are fatal for many men; and on the mental side, there is the simple fact that girls, on the average, do better at school than boys."

Because the instructions preceding the argument called for criticism of the position taken, students understood that they were to find something wrong with the argument. Many students simply attacked the basic assertion that "Women are superior to men," as if the question had been worded that way. A large number,

however, responded to the test-maker's intention and attacked the premises of the argument, as well as agreeing or disagreeing with the central assertion and offering some defense of their own views. Scoring these papers involved the tester and the readers in some controversy over whether students who addressed the assertion without attacking the argument's faulty premises had adequately addressed the test question. Ultimately, readers insisted on giving full credit to students who interpreted the instructions as an invitation to quarrel with the position stated, as well as to those who pointed out deficiencies in the form of the argument, since the former set of papers included some of the best writing in the sample. Actually, the disagreement between test-maker and readers reflected a deeper disagreement about the purpose of the assessment: whether the measure was to be an overall test of self-expression or a more precise test of certain reasoning abilities.

The scoring of this test question could be based on whichever of the two purposes were paramount, but in either case, it revealed both the dangers and the advantages of using highly controversial topics. All students wrote fairly fluently to the topic; many wrote with passion and conviction, drawing on personal experience, assessing personal values, qualifying definitions (of "superior", for instance) or discussing the practical implications of taking the given position. Even poor writers generally made one or more cogent points. On the other hand, some students inevitably expressed views that offended some readers. One boy was scored down by two readers (one male, one female) because of the extremism of his views. Although he had intelligently attacked the faulty premises of the given argument, he had frankly admitted to biases that the readers thought outrageous and offensive. In a test of general writing ability, his low score was misleading, because he had written cleverly and effectively.

Here the holistic nature of the judgment breaks down, when the readers are to reconcile the quality of the writing with the presumed wrongheadedness of the ideas expressed, and they feel compelled to judge one to the exclusion of the other. Such papers often receive discrepant scores in their two readings.

Controversial issues can be presented interestingly as legitimate conflicts of interest, as in the following prompt which was designed to test powers of observation, reasoning, and argument through a persuasive writing task administered to a group of entering freshmen in a large state university (Cooper, et

al., 1979).

PROMPT 31a

At the place where you work, a woman has just quit her job, leaving vacant the company's only executive position ever held by a female. The Board of Directors has stated their preference that a woman replace her in order to fulfill an Affirmative Action quota. As a member of the Hiring Committee, it is your job to help choose a successor to the post.

The only woman who has applied for the job seems competent and meets the written qualifications for the job but she is clearly less qualified than both of the men she is competing with. Members of the Committee disagree about what should be done: some say hiring a woman is absolutely necessary for breaking down employment discrimination; others say hiring a less qualified person would be foolish as well as unfair to those working under the new executive.

To have a full hearing of all views on this critical issue, the Hiring Committee has asked each member to prepare a carefully written statement to be distributed in advance of a meeting to discuss the issue. Write a statement which represents your position in the matter, making it as logical and persuasive as possible. Your writing task is to persuade the Committee to adopt your own view and to vote on the job candidates in accordance with your view.

Note that in this prompt, the underlying topic could be described as the issue of affirmative action employment policies. The prompt might have been offered in this form:

PROMPT 31b

Argue pro or con affirmative action policies in admissions or employment.

Instead, the task was set in the form of a simulated life-situation in which the writer's attitudes toward affirmative action were to be expressed in the context of recommending action. Actually, this prompt can be

criticized for being somewhat loaded: The phrase, "is clearly less qualified than both of the men she is competing with," led some students to assume that the tester thought the female applicant should not be hired. A less judgmental wording would have been: "is less experienced than...."

Sometimes even pilot-testing in advance fails to reveal weaknesses in persuasive writing tasks and these weaknesses show up later in the testing of large populations. Of three prompts used in the 1970's on a statewide proficiency test,<sup>8</sup> only one met the important criteria for persuasive prompts: that they a) be relevant (i.e. students should care about the issue); b) draw on an adequate information base (i.e. students should know something about the issue); and c) be non-biasing (i.e. different positions on the issue may be defended reasonably).

On this statewide proficiency test, over three different years, students were asked to take a position on these issues: a) whether physical education should be a required course in high school, b) whether firemen should be allowed to go on strike, and c) whether students should be required to produce a writing sample as part of the proficiency test. According to a number of teacher-readers, the first issue generated wide appeal and students generally were able to advance valid arguments on both sides of the question. On the second issue teacher-readers reported that fewer students had strong opinions or even understood the underlying issues in the debate. The third issue proved to be relevant but somewhat loaded; that is, teacher-readers reported that no arguments against the writing sample requirement appeared reasonable to them. One teacher-reader believed that the prompt was culturally biased, claiming that students who argued against the writing sample requirement came from particular ethnic groups that did not share mainstream emphasis on the importance of writing. Being unfamiliar with the arguments for using a writing sample to measure proficiency, these students were unable to provide convincing counter-arguments for the teacher-readers, all of whom were, to some extent, biased in favor of the writing sample requirement.

In short, as with imaginary situations that can become "loaded questions," problems result when students are asked to take a pro or con position on a

---

(<sup>8</sup>CHSPE, the California High School Proficiency Examination, administered by the Educational Testing Service)

controversial issue on which only one position is considered defensible by most or all of the teacher-readers. Problems also result if a carefully qualified position that is neither pro nor con represents a more intelligent response. In these cases, students may feel that the prompt does not allow them to straddle the issue, and so they will adopt an artificially simplistic pro or con view, thus failing to write either convincingly or interestingly.

### Topics With Hidden Biases

In attempting to guide students towards more fruitful ideas, a test-maker will sometimes load a topic with words that unintentionally suggest the test-maker's biases. Students sensing the bias may then write from their conception of what the test-maker believes rather than from their own perceptions and opinions.

The following topic was pilot-tested at a state university as part of an equivalency examination:

#### PROMPT 32

Consider the words trees and plastic. Write a paragraph in which you discuss some aspect(s) of what the words suggest to you.

Students tended to fall into one of two kinds of traps in their responses. Some students simply catalogued cliches about man's abuse of the environment, using "trees" and "plastic" to represent a kind of polarization of the natural versus the man-made. Others, attempting to go beyond cliches and generate an original thesis, simply got bogged down, and the resultant writing was strained and even incoherent.

The prompt was revised to read simply trees. In response to the new prompt, students produced interesting essays displaying an enormous range of reactions to the word "trees", a range that included discussions of family trees, branching tree diagrams, wilderness treks, the uses of wood, and man's encroachment on nature. The single word clearly elicited what the original topic was meant to inspire. (Plastic was later used alone with excellent results as well.)

Finally, a kind of unintentional bias occurs when the test-maker assumes that students share certain attitudes or that they perceive themselves as the test-maker perceives them. When the following prompt was piloted for a college proficiency test in 1980, the test committee was certain that the prompt would have

180

universal appeal for students. Their certainty was based on the experiences of their own late adolescence during the turbulent, demonstration-filled 1960s and 70s. They assumed that even though marches and protests now seem to be less widespread, all young people perceive themselves as rebels at some time:

#### PROMPT 33

Even those of us who may conform outwardly have our private, personal forms of rebellion. Some rebel against society's conventions, some against family values, and others against the attitudes of peers. We might dress in unconventional ways, reject our family's traditions, or disagree with friends' political beliefs. Some of us may simply rebel against popular fads. Many of us are not interested in jogging five miles a day, eating natural foods, or wearing designer jeans.

#### INSTRUCTIONS:

Describe one way you choose not to conform. Explain why you act as you do and analyze the effect your choice has on your life.

On the whole students had little to say in response to this prompt. Few had actually experienced "personal forms of rebellion" which they took seriously and thus could write about convincingly. In reviewing the pilot essays, the test committee felt that this group of students was less concerned with rebelling than with conforming, that they could not identify social values, mores, or fashions which they wanted to rebel against in order to establish their own identities; rather, these students saw themselves as fitting into the present life-styles of friends and families. On the basis of the pilot-testing, the test committee chose an alternate prompt, one which asked students to write about making a difficult decision. The response to this prompt was excellent; it had tapped a core of experience important to this particular group of student-writers.

To avoid biasing effects in either writers or readers, test-makers need not abandon thought-provoking questions that allow students to explore values or investigate ideas. When evaluators wish to test the reasoning skills needed for exposition or argument,

they need not avoid using challenging statements as prompts for writing. The following prompts present generalizations which can be supported or challenged in a coherent essay, enabling the student to demonstrate many of the skills needed for academic writing, while allowing them to draw on personal experiences and express personal values:

PROMPT 34

Man is a creature of habit.

Do our habits make us more stable, more efficient, or just duller? Discuss. Support your arguments with one or two specific illustrations from your reading or experience.

PROMPT 35

Opposition is indispensable. A person will often learn more from the people who disagree with him or her than from those who agree.

Do you agree with this observation? Have you profited more from people whose views were similar to yours or from people whose views opposed your own? Discuss, using one or two specific examples to support your position.

These two prompts invite an exploration of ideas which need not be argumentative or persuasive in tone or purpose. Further, these prompts require a thoughtful development that makes them appeal more to college-bound or older students. The following two prompts--neither dealing with current controversies--elicited strongly persuasive writing from ninth through twelfth graders at all ability levels in a Bay Area high school:

PROMPT 36

Not all inventions have been good for humanity. Name one invention we would be better off without, and make it clear why. You may do this writing as an essay, journal, letter, story, or other form.

PROMPT 37

Imagine that a small group of people will be sent to colonize a new planet. Food, clothing, shelter and transportation have been provided for. You are among those asked to select a few additional things to be sent along in the limited space available in the ship. What one item would you recommend, and why? You may write your recommendation in the form of a story, a dialogue, a letter, a speech, an essay or other form.

The responses to the invention prompt ranged from serious political or social arguments to humorous proposals suggesting abolition of such dreaded inventions as the waffle-iron or hair curlers. The objects recommended for the colonizing of the new planet included everything from important books intended to help the new community avoid further war to a jar of peanut butter, proposed seriously in order to inspire curiosity and comment among the new colonists and thereby ensure a certain degree of levity in all future decision-making.

## THE WRITING INSTRUCTIONS: PURPOSES AND PROBLEMS

### Section Overview

As described above, the writing prompt has two elements--the topic which states the required content of the writing (the what) and the instructions which give the suggestion about or limitations on the content of the writing (the how). This second element can also be called simply "the writing instructions." The writing instructions may be very simple, consisting merely of the phrase, "Write about..." or they may be quite elaborate as the test-maker tries to guide the writers in their responses. As with cuing for content, cuing for form can help assure that students understand what is expected of them. The prompt-writer's dilemma is to provide enough guidance without providing too much. This final section considers the following questions relating to such instructions: a) Do they clarify the test-maker's intention, leading students to respond in ways the scorers will reward? b) Do they effectively limit what the writers attempt, making writing samples comparable? c) Are the task restrictions helpful, or are they likely to distract students from the central task of communicating their ideas in writing?

### Hidden Agendas in Instruction

The hidden agenda on a writing test is the set of implicit expectations that every test-maker has about how the student-writer should respond to the prompt. Students often understand these expectations but sometimes an unintended trap is created. Pilot testing of prompts will usually reveal cases in which the test-maker's assumptions about what the students will understand prove erroneous, requiring more explicit directions. Sometimes what is needed is not a new prompt but simply a willingness on the part of the teacher-readers to accept a broader range of possible responses than were originally anticipated.

There are three common traps, however, that might be avoided by a careful search for hidden agendas:

#### The "Describe" Trap

An interesting photograph, previously used successfully as a writing stimulus, was presented as a test prompt to two classes accompanied by the instructions "Describe what you see in the picture above." One of the two classes had been accustomed to writing stories and essays based on pictorial stimuli. They had been practicing making accurate observations and writing descriptions by writing simulations of police

and insurance reports. The second class had not engaged in similar work. The students in the class with the specialized practice in description wrote the barest, most literal descriptions of what they saw in the test prompt photograph, refusing to make many inferences or interpretations. These students received low scores from the tester because their writing was the least interesting and imaginative. The tester's hidden agenda involved a definition of "description" for this task which required writers to be imaginative, not literal, and make inferences about what might be happening and about the mood inspired by the picture.

The "describe" instruction has been problematic in other assessments. One school system found that the task, "Describe a person who is important to you," elicited from many writers the most limited physical descriptions, while a revised instruction, "Tell about a person who is important to you," seemed to suggest to students that they could select whatever facts about their subject interested them and might interest a reader. Some teachers argued that the "describe" version of the task was a good test of whether students knew what it meant to describe and that description of a person can be much more than a list of bare physical characteristics. But in this case, the instruction to "describe" again made the test more a measure of the students' ability to recognize the test-maker's meaning rather than a measure of their own ability to write interestingly about a person important to them.

In general, it can be said that any writing instruction consisting only of the simple direction, "describe," is a case of underprompting and creates an unfair hidden agenda unless cues are added to suggest what the tester desires to see. Prompts such as, "Describe your favorite person," "Describe your favorite object," "Describe your favorite food," are popular topics with students, and can be made more indicative of the tester's expectations by adding a qualification such as the one illustrated by the underlining:

#### PROMPT 38

"Describe your favorite \_\_\_\_\_ (person, food, object) so that your reader will understand why it is your favorite."

This simple addition clarifies the rhetorical purpose for the writers and cues them to recognize that the test-maker is interested in more than merely the ability to write accurate physical descriptions.

Other casualties of underprompting from using only the unmodified instruction, "describe", occurred in two tests asking students to describe places: "Describe your home" and "Describe your school." The problem with these prompts is that they are devoid of rhetorical purpose. Thus, the descriptive task becomes overwhelming without the limiting instruction to choose a favorite something and explain one's preference to someone. The students, in effect, must set their own limits and construct their own rhetorical purposes as follows: Why am I describing my house or my school? Who wants to know what? How can I possibly describe all of it--and, if not all, then what aspects should I attempt?

Some students, familiar with such vague tasks, do quickly impose their own rhetorical purposes and describe homes or schools, evoking feelings about these places. Most students, however, painfully and conscientiously begin a "guided tour of their home" or try to portray graphically the outside of the building, giving facts about its location, its height, its color, etc. In this case, providing the additional instruction: "Describe...so that your reader can understand what it looks like" isn't very helpful. The writers are still restricted to providing a physical description of a large, complex structure that they are not looking at. They would still not be sure whether their examiners expected them to convey feelings and attitudes or "just the facts."

It is interesting that one major research study foundered when an entire group of subjects in eleventh grade simply refused to write on the following prompt because they felt it was too "childish" (Stahl, 1977).

#### PROMPT 39

My Home: Write a description of your home, its rooms and their contents in such a way that someone who has not visited it can form an idea of the way it looks.

What the researcher did not realize is that it was also too difficult, that only the most mature and confident writers can take such an undeveloped subject and provide their own point of view, their own definition of the task, their own rhetorical purpose and thereby create a coherent and meaningful piece of writing.

A sophisticated version of the "describe an object" prompt with detailed instructions has elicited good expository writing in state college writing assessments conducted for the purpose of gathering

information for making admissions decisions (White, 1973):

PROMPT 40

Many observers of our society claim that modern man, immersed in materialism, is "owned by his objects." Yet many of us have objects that we treasure not just for their material value but for a variety of other reasons. Assignment: Describe one or more objects which are important to you. Explain what values they represent and comment on those values.

Walker Gibson (1959), in Seeing and Writing, provides a class assignment to describe a church:

PROMPT 41

Find a church in your neighborhood that seems to you an interesting building. Describe what you see on its facade in detail and in such a style that your reader sees it as interesting too.

He goes on to explain what a complex task this apparently simple description is, and why a church is a good subject for this kind of practice. He suggests that churches are often loaded with personal meanings, and that the task entails working out "who you are, who you should be, as you speak to a reader about the appearance of a church." Description must be done with both a purpose in mind and a point of view in order to re-create the actual experience of looking. Description is a sophisticated exercise even for advanced writers; it is not, as might be expected, the simplest mode of writing and therefore it must be used judiciously in testing for general fluency. (Gibson's church description assignment appears as Theme 8 in his course sequence of 15 exercises.)

Even when the goal of assessment is primarily to test students' ability to develop an objective description rather than to express personal meanings through description, the instruction, "describe", can still be a case of underprompting:

PROMPT 42a

Describe what you see. (Accompanying a photo and diagram of an automobile accident.)

Some students will write about the horror of the accident, others will try to infer who is to blame or will write about what must have happened, while a few will try to give a precise description of the position of cars and bodies. If test-makers wish to elicit this last kind of writing, their instructions must be more explicit, specifying a rhetorical purpose:

#### PROMPT 42b

Study the accident in the photo, using the diagram to help you gain a clear understanding of the situation. Describe the accident, reporting details accurately so that a jury or insurance company can make a fair decision about who is to blame.

This task is a powerful test of many skills that contribute to writing ability, including the ability to marshal facts and observe details in a reasonable way so that a reader can grasp the important aspects of a situation--a valuable skill worth testing.

Such tests can be scored holistically as well as analytically, of course. But student writers must share a common understanding of the rhetorical problem they are trying to solve if teacher-readers are to make fair comparisons of their solutions.

#### The Special Conventions Trap

It is, of course, not a trap to test students for their knowledge of special conventions, such as the forms of business letters or friendly letters, as long as the test-maker does so deliberately, deciding in advance which forms are acceptable and how deviations are to be scored. Teachers and students must be informed of the conventions which will be tested in the writing assessment. The trap is present only when certain conventions are sprung upon unwary teachers and students.

In an effort to create a meaningful rhetorical context, test-makers may call for students to write business letters, but fail to decide in advance how to weight formal considerations such as salutations, closings, return addresses, etc. Such a problem arose with this otherwise excellent writing prompt.

#### PROMPT 43

The following advertisement appeared in the local paper:

## SUMMER EMPLOYMENT

### Earn & Learn

On-the-job training for future positions in:

Medical Services Aide at Highland Hospital  
Ticket taker at local theatre  
Grocery clerk at local supermarket  
Plumbing Assistant for plumbing company  
Assistant to the manager of a circus

Earn \$3.50 or more per hour while you learn a valuable skill.

Send letter of application to:  
Ms. Laura Jones, Opportunity Training Center  
2212 Fruitvale Ave., Oakland, California  
94231

Pretend that you are applying for one of the jobs listed. Write a letter explaining why you should be hired. You should give information about yourself. The information may be actual facts, made-up information, or a combination of fact and fiction. The letter should help you get the job.

Students wrote lively and interesting prose, using real or imaginary facts. But they also used a variety of different conventions in setting up their letters. Teacher readers differed in their preferred conventions; some believed that "correct form" called for no paragraph indentation. Still others stressed the importance of adequate spacing between the date and the body of the letter.

If all students have taken a course in business English or typing and have been taught the same set of conventions for business letters, it would be possible to devise some system for scoring the finer points of layout and presentation. However, it is clear that holistic scoring would not be the most efficient or accurate way of measuring mastery of these conventions. A rating of letter form might be an appropriate addition after the holistic assessment has been made.

In the assessment for which this prompt was designed, testers were not concerned with conventions or consistent advance instruction to students in the fine points of the business letter. Rather, their purpose was to determine whether students could express themselves adequately in a life-like situation.

Assuming that students might refer to one of several business letter guides should they actually wish to write a letter applying for a job, the teachers' greater concern was whether students could include appropriate information about themselves and write in a convincing, personable style that would appeal to prospective employers. Obviously students needed some control of punctuation and spelling, of vocabulary and sentence structure, if they were to sound mature and reasonably well-educated. But in the holistic, rapid-impression scoring, these factors could easily and fairly be weighted for what they contributed to the over-all effectiveness of the letter. Since no one was really applying for a job as a business secretary, the use of particular conventions would, even in real life, be less important than the overall impression conveyed by the writer. In these circumstances, teacher-readers agreed to suspend debate on formal conventions and rule out of order any penalties imposed for incorrect letter etiquette.

When offering such topics, testers should note that in the absence of specific directions to attend to conventions, students may differ in their assumptions about what the tester wants. Students understand that this is an imaginary letter, not a real one. Some of the best students may, under the circumstances, drop all but the barest suggestion of a letter: "Dear Sir," and "Sincerely," being the extent of their formalities. Other students may pay close attention to letter forms assuming that this is what is being tested, and may offer only a minimum amount of content, as in this example of impeccable form:

WRITING SAMPLE:

1204 Kenyon Road  
Oakland, California 94724  
March 2, 1978

Ms. Laura Jones, Director  
Opportunity Training Center  
2212 Fruitvale Avenue  
Oakland, California 94321

Dear Ms. Jones:

I am very interested in applying for the position of Medical Services Aide at Highland Hospital, which was advertised in the OAKLAND TRIBUNE last Saturday.

I am sixteen years old, and have had experience baby sitting, attending old people, and volunteering as a Candy Striper. I have excellent references, which I have enclosed on a separate sheet. I am available for an interview any day after 2:00. My phone number is 634-0972.

Sincerely yours,

Pat Smith

No one could find fault with the formal aspects of this letter of application; however some teacher-readers felt it suffered by comparison with the letters of students determined to win the job in this one communication act without recourse to references or interviews. Some of the fuller, more personal of these letters were strong, impressive pieces of writing, although strictly speaking they revealed less awareness of appropriate business formalities than the letter above.

Teachers may justify teaching conventional organizational formulas, such as the five-paragraph essay, believing that they are providing students with structures to order their ideas for greater readability. But the evaluators are not justified in penalizing students for not using those forms in an assessment of composing abilities. Instead, the major concern is

whether students can achieve their rhetorical purposes by whatever means they choose. Holistic assessment puts reader and writer into a naturalistic relationship, one in which the reader is allowed to respond not as an error-counter or a conservator of threatened forms, but as a receiver of intended communication.

### The Pick-a-Form Trap

Sometimes a prompt will attempt to encourage students to experiment with fictional forms or to allow them maximum freedom to match form to intention:

#### INSTRUCTIONS:

Write in whatever form you want--journal, letter, essay, story.

This kind of instruction with a range of options gives opportunity for the consideration of techniques which may be in the process of being taught in the writing program. Thus, imaginative students are allowed a full range of choices for creative experimentation which often results in outstanding writing. Further, the use of such a test instruction encourages teachers to provide opportunities in class for students to practice a variety of forms.

But the apparent freedom of this instruction may become a trap if the readers who score these papers have a "hidden agenda" that rewards conventional school-essay language. Students who choose to write letters or journals may imitate the speech of their characters and write less formally or correctly than they would have if they had chosen more conventional essay or story forms. The tester must be prepared to train teacher-readers to score experimental papers so that the student is not penalized for attempting the very fictional techniques that have been suggested in the writing instructions.

A wide choice of genres can have another unfortunate effect. Even when teacher-readers are conscientious about giving full credit to forms other than traditional essays, they may find it difficult to compare papers of different genres. It is easier to judge the relative merits of two short stories than it is to judge the relative merits of a short story compared with an argument. Many experts in writing assessment warn against providing a wide range of forms because it is likely to decrease inter-rater agreement. When the purpose of the assessment is to encourage the teaching of writing for different purposes, or writing with

different forms, administrators of assessments should consider offering several writing tasks, each clearly calling for a different mode of writing.

In practice, however, the given topic somewhat limits the genre and rhetorical strategy choices, so that papers in a particular year can be scored quite reliably, even though responses range across a variety of forms. In the case described above (in "Hypothetical Situations As Topics"), the option to select forms actually helped students in certain years to solve special problems inherent in that year's topic.

One further warning relates to poetry as a form of response. Many schools do not have large numbers of teachers who could confidently evaluate and score poetry; even poetry teachers find it hard to rank order poems in comparison with prose works. Therefore, schools find it best for purposes of general assessment to exclude poetry as a possible mode of response.

#### Helpful and Unhelpful Instructions:

Suggestions That Become Constraints. Sometimes in an attempt to avoid the hidden agenda problem, test-makers strive to make explicit every criterion the writer should know about, to provide every conceivable cue that might elicit a good response. If teacher-readers are expected to reward abundant use of details and personal feelings, the test-maker might write:

#### INSTRUCTIONS:

Be sure to include lots of details and to make your feelings clear.

For the writer, however, what is intended as a suggestion is transformed inescapably into a constraint which may or may not be helpful. An instruction that is intended to provoke more richly textured writing, "Use lots of details," may imply that the students who do not will be penalized. The instructions "Support your opinion with specific examples," or "Write a well-organized essay," assume a certain knowledge of structure on the students' part and serve as reminders to them to use that knowledge as they compose. Such instructions are expected to direct their attention to what it means to write a well-organized essay, to consider the use of illustrative anecdotes rather than unsupported arguments or ill-formed generalizations. For many students these cues, coming in the context of a school setting in which they have practiced responding to similar instructions, can be helpful. Since

they trigger memories of other school tasks which have been evaluated, such instructions may provide students with some idea of how this particular task will be judged.

But two things can turn a helpful cue into a damaging constraint. Students' concern with doing as the tester asks may keep them from attending to their own construction of meaning in response to the topic, and this can cause them to produce strained, awkward essays rather than the fluent pieces of which they may be capable. Or students who attend to the spirit rather than the letter of the topic may produce outstanding responses, but be scored down because the scoring guide specifies that top scores should be awarded only to students who solve the problem in the way specified. The following constraints are likely to divert the student writers' attention from communication to secondary matters and so prevent them from accomplishing the primary writing task.

Effects of Overprompting. Attempting to make up for the deficiency of a test prompt such as "Describe your school" or "Tell what you see in the picture," many test-makers add a stream of instructions and suggestions which sometimes have counter-productive effects on the writer. Three examples demonstrate these elaboration techniques which tend not only to turn the writing test into a reading test, but to burden the writer with choices which may prevent simple and natural responses.

One tester sought to avoid the underprompting problem of "Describe your school," by listing aspects of the school which might be described.

#### PROMPT 44

Describe your school. Tell how it looks from the outside. Tell how your classroom looks. You may describe what happens in the hallways at different times of day, or you may describe the lunchroom or playground. If you like, you may describe the people in your school.

This type of overprompting appears when testers seek to insure that writers supply adequate details. However, this elaborated cuing encourages an unfortunate tendency in many young writers to describe superficially many different things rather than to concentrate on one aspect of a subject long enough to write effectively. The additional cues might not be so destructive of coherence and purpose if they were reworded to

encourage students to focus on one part of the school (See the discussion under "Focusing," below). As Prompt 44 is worded, students who begin conscientiously to "tell how it looks from the outside" may rarely get any further than that. They might write much more interestingly if they had chosen to begin to describe the playground or the lunchroom. The student might also be more helpfully guided if the instructions were rewritten to invite them to describe one part of the school in a way that would be helpful to a new student who is about to enroll there.

A second kind of overprompting can result when testers wish to insure that they will get a particular kind of writing, knowing that students may be misled by such a prompt as "Describe what you see in the picture." The National Assessment of Educational Progress moved away from holistic assessment toward primary trait scoring precisely because the test-makers could not predict that all students would, for example, narrative or to write imaginatively in response to a pictorial stimulus. Invariably, such stimuli would inspire some students to write discursively or factually.

In their 1973-74 assessment, the NAEP offered writers a photograph of children playing on a boat. The instructions read:

#### PROMPT 45

Look carefully at the picture.

These kids are having fun jumping on the overturned boat. Imagine you are one of the children in the picture. Or if you wish, imagine that you are someone standing nearby watching the children. Tell what is going on as he or she would tell it. Write as if you were telling this to a good friend, in a way that expresses strong feelings. Help your friend feel the experience too. Space is provided on the next three pages.

When the researcher tested this prompt on a group of teachers, one of them complained that it sounded like the old jibe, "Keep your ear to the ground, your nose to the wheel, and your shoulder to the grindstone: now try to work in that position." Perhaps the prompt can be justified for NAEP's purposes, since they are testing students from enormously varied backgrounds for a particular ability (i.e. the ability to enter into the imaginary world of the picture). But this prompt makes many demands on the writer which may be at odds

with the writer's spontaneous response to the picture. Simple, relatively open instructions, like those accompanying a picture stimulus in the Jefferson County, Colorado, writing assessments (1976 & 1977) should be considered:

PROMPT 46a

Look at the picture above. You can react to it in any way you wish by writing about it in one of the following ways: as a story (What have the characters been doing? What could happen next?), as a description, or as an opinion.

If testers wished to rule out the possibility of mere factual description, which is likely to produce writing that is less interesting than stories or essays, the instructions could read instead:

PROMPT 46b

Look at the picture above. Let it give you ideas or create a mood. Then use the ideas or the mood to write a story, a memory of something that happened to you, or an essay or argument based on some strong feeling inspired by the picture.

Since this kind of instruction opens up a large number of choices for writers, they must be provided with adequate time to choose their point of view and mode of response.\*

A third type of overprompting can occur when testers wish simply to remind students about aspects of their training which should be applied to the current task. The following prompt used in a district assessment illustrates this kind of overprompting:

---

\*The NAEP prompt attempts to provide a single mode and point of view partly to help the writer begin to write more quickly, but primarily to limit the range of likely responses. These goals in themselves may be suspect if the writing test is an attempt to measure how students actually compose when faced with an opportunity to write about an interesting stimulus.

PROMPT 47

"No man is an island, entire of himself.  
Every man is a piece of the Continent, a part  
of the main..."

John Donne

This topic should lead you to think about the possible meaning of the quotation. You may build upon the idea by supporting it, or you may contradict the topic with an opposing point of view. Attempt to build a logical, reasonable opinion paper. Include an introductory and concluding paragraph and two or three paragraphs in the body of the paper. The supporting paragraphs may be developed by comparison/contrast, definition, illustration, or any other form which will produce a unified essay on this topic.

This prompt was pilot tested and used successfully by conventional measurement standards: it elicited papers that holistic scorers were able to distribute along a full range of points from best to worst. But the extent to which students benefited from these elaborate instructions is unclear. Current research on topic effects promises to provide insight into what information students use when they approach assessment prompts.<sup>8</sup>

Specifying Number of Words, Sentences, Paragraphs.

In most writing assessments, test-makers need to impose a time constraint which indirectly controls the length of student work. Some assessors, in addition, impose a length constraint in the form of a minimum number of words, sentences, or paragraphs. It is not uncommon to see:

INSTRUCTIONS:

Write at least three sentences... (elementary)  
Write one paragraph describing... (elementary)  
Write two paragraphs in which you...(junior high)  
Write an essay of about 300 words...(high school,  
college)

---

<sup>8</sup>See Chapters 2, 3, 4 and 5.

In general such constraints are unhelpful, both to students and testers. The testers' purposes can be served in other ways that do not interfere with the primary aim of the testing, which is to observe how much and how well a student can write in response to a prompt.

The instruction in the elementary school to write three sentences is intended to insure that the tester can determine the extent to which children understand what a sentence is and how sentences are marked. This aim is more appropriate to objective testing in which testers can measure sentence recognition, punctuation skills, etc., as skills separable from the act of composing. But, in an actual trial, because the prompt asked for three sentences, several third grade students whose sentence combining skills outstripped their peers were penalized for writing one, long, layered sentence which included more information than appeared in the three-sentence papers. Other students could not think of three sentences on the given topic ("Describe a hamburger"), and so they added a sentence about some other kind of food (such as "chocolate cake" or "hot dogs") and were penalized for being off topic. One student even construed the instruction to mean that she should copy her one sentence over three times. Most of the third graders in the study simply ignored the three-sentence constraint and wrote one to six sentences of varying quality about hamburgers. The majority of students wrote two sentences.

A more likely way to insure that students will generate more than one sentence is to give them an interesting rhetorical problem--one which provides some context to the request for a description of a hamburger. In any case, one of the most important things a tester needs to measure in elementary school writing is the degree of fluency--the number of sentences or ideas the child can generate when given a stimulating topic to write on. For children who can only think of one sentence to write, the command to write three sentences is simply an assurance of failure before they begin--if they even notice the instruction. These low fluency students will be identified whether or not the constraint on length is included in the test instructions. In a test the instruction to write a set number of sentences on a topic not only may encourage redundancy, it can increase writing anxiety and turn students' attention to arbitrary formulas for composing rather than help them generate ideas.

The problem in specifying "Write one paragraph" is somewhat different. If the focus of teaching has been the construction of paragraphs, and if students have been practicing this unit of discourse, a tester may

well justify a test which gives them a familiar instruction and tests for paragraph recognition and construction. But the virtues of this approach both to teaching and to testing are debatable. Some students will already be sufficiently fluent to produce material for two or more paragraphs even before a teacher introduces the class to the concept of paragraph construction. Others will still be struggling to generate enough material to elaborate a single idea. Teaching the paragraph as a complete unit of discourse rather than as a method of shaping the flow of ongoing discourse is misleading, even at early stages of writing instruction. More fluent students can be taught to indent, breaking up longer pieces, as an aid to readability. Less fluent students can be taught to elaborate, and can be helped to understand some of the questions that readers wanting to know more about their subject might ask.

On any test of general fluency, the crucial variables will be whether students are able to write fluent, coherent discourse that communicates their intentions. Performance will depend on response to the topic, and paragraphing will depend on the extent to which they have learned to use the paragraph to structure their thinking or aid their readers. The instruction "Write one paragraph..." or "Write a paragraph..." introduces a concern for the paragraph as a unit of testing rather than as a unit of communication. It leads a student to think: "What kind of paragraph does my teacher want? Suppose what I want to say doesn't fit into one paragraph? Should I change my idea, go ahead and write two paragraphs, or just skip the indentation?"

Again, as with specifying a certain number of sentences, if testers' intentions include measuring the ability to recognize where to indent or the student's sense of the internal coherence of a good paragraph, they can develop reasonably good multiple choice items. Using a writing sample to check primarily on "paragraph skills" is a case of using the wrong testing format. The writing test should allow students to demonstrate control of paragraphs as naturally-- occurring units in their composing process.

Finally, the constraint for older students--to write 100-500 words--whether included in a test instruction or on classroom assignments, encourages pointless word counting that typically results in padding rather than in real idea development. Where the word count is intended to keep students from writing too much, or to encourage conciseness, the effort to condense, is--among experienced writers--typically a

revision activity, and a complex one at that. Most test-makers who include the word-count restriction do not intend a close counting but merely mean to suggest the general scope of the topic--how much students should attempt to do in the time available. The word limit presupposes that students have some idea of what a 300-word paper is--like newspaper columnists who write regularly to a given format. A more natural limit on what students attempt to write is imposed by the amount of time provided. Adding any other constraint on length may lead the writer to perceive word limit as being in conflict or incompatible with the time constraint.

Although this informal observation about the effects of specifying number of words remains to be tested in careful research, students should be given enough time to deal with the prompt. The number of words would then be determined by such relevant considerations as how much they find to say and how much they are able to revise, in the time available. If they are not writing for publication, the number of words attempted is probably the least important aspect of the composing task. In a testing context students should be using their energies and time on concerns other than counting words.

#### The Value of Focusing Cues.

One of the most helpful constraints that can be provided in writing instructions is a suggestion to focus on one aspect of a larger topic, or on one problem, or on one or two examples in support of a position. When an evaluator tested the following prompt, the instructions elicited long lists of general complaints with little analysis of problems, few suggestions for solutions:

#### PROMPT 48a

Write a letter to your principal. Explain to him or her what you would like to see changed in your school.

The tester had expected students to write as if they were really trying to bring about change. The students, on the other hand, interpreted the task as a general sort of complaint letter--a "gripe list"--for anyone who will listen. A revision of the prompt with a focusing constraint provided in the instructions resulted in more fully-developed arguments:

#### PROMPT 48b

There are probably several things you would like to see changed in your school. CHOOSE ONE OF THESE and write a letter to your principal, describing or explaining the problem and suggesting what he or she might do to change the situation.

Another example of the value of providing helpful focusing cues appeared during the development of a College Board English Composition Test prompt. One year students were given a provocative quotation and instructed to write about it, drawing on examples from their experience or their reading. The prompt generated writing that represented the full range of ability in the population tested when it was administered and scored. The following year, however, when the instructions read "Choose one or two examples from your experience or reading," one reader compared results and noted clear improvement. The earlier version had elicited writing that consisted of loosely listed anecdotes. The latter version tended to elicit decidedly more coherent writing, with students focusing their efforts on single, telling examples, or drawing on one or two pieces of literature in some depth.

One word of caution in using this apparently helpful strategy: having cued students to concentrate on one or two ideas, testers should leave room in scoring for the exceptional student who, working at a different level of abstraction, may refer to many examples, but does so convincingly, without sacrificing coherence or depth.

#### The Effects of Specifying Audience.

Specifying audience--that is, providing writers with an imaginary audience other than the test reader--may be useful in some testing situations, but it may cause problems in others. When audience is specified in a writing task that provides a full rhetorical context and a clear purpose for writing, the effect is to create a more realistic writing problem, as in the following prompt for elementary and middle school or high school.

#### PROMPT 49

The following article appeared in the local newspaper:

KDOL Begins Survey:

Why do we remember certain people from our

past and forget others? Television station KDOL is preparing a special TV show on personalities people remember from their elementary school years--a classmate, a member of the school staff, a family member, or a neighborhood friend. Send your description to:

Program Producer  
KDOL  
Oakland, CA 94606

Respond to the article by choosing someone who made an unusually strong impression on you. Write about that person, explaining why you remember him or her. Do your best work.

Testers seem to feel that merely mentioning an imaginary audience in a prompt will help students write with more enthusiasm and purpose in a writing exercise that is still a school test, one whose real audience is ultimately the teacher-scorer. A real danger lies in the injudicious selection of an audience cue--for example, "Write for a friend." Following the lead of this instruction, some students will conscientiously adopt a tone more appropriate to writing for peers than for teachers. Because this language is less formal, often less articulate, and sometimes less comprehensible to readers than the student's "school language," the student who is mature enough to vary language for different audiences may actually be penalized during the holistic scoring.\*

In the one of the studies conducted under this BAWP/NIE Writing Assessment Project, the prompt had invited students to think of a first-time experience and to write about this experience for someone who is about to have a similar experience.\*\* Student-raters of the papers said that certain ones sounded like English compositions rather than like the friendly letters they pretended to be. These "English composition" types were marked down by students, but they received top scores from teacher-raters, while other papers that received low scores from teacher-raters received high ratings by students because of their verisimilitude.

\*Marion Crowhurst (1978) has researched the differences in student language when writing for peers and writing for teachers on a school task. In some modes, students write less formally, using syntactically less mature language when assigned to write "for a friend."

\*\*See Chapter 2, Part II.

Trial responses should be examined closely for signs that students are using less mature, less explicit, less effective language than they might if they were addressing their actual audience of unknown adults.

Other effects of specifying audience may surprise the unwary tester. One teacher reports that he has stopped asking students to write to pen pals because the students seemed to have more to say to other types of audiences: actors, television or radio characters, authors of books and articles read in class, members of their family, friends they have missed since their last move, makers of products that they liked or didn't like, other students in the school. Consequently, when this same teacher was asked to administer the following prompt as part of a district assessment, his students complained that it was "dumb," "boring," or "phony," and wrote very little, giving a misleading impression of their writing abilities:

PROMPT 50

Write a friendly letter to a pen pal in another country, telling him or her about your home or school.

But students in other classes who had had little experience with writing for an audience did not protest the prompt. In this case, the teacher's own class required a more sophisticated choice of audience because of the versatility developed during prior writing assignments.

Other attempts to specify a real audience also have been problematic. For instance, some assessments may be successfully using one or another version of the "letter to the principal" prompt discussed earlier.

PROMPT 48c

There is probably at least one thing about your school that you would like to see changed. Write a letter to your principal and tell him about the problem. Be sure to describe the situation as it is now and tell why it bothers you, as well as how you would like to change it.

This version of the prompt, however, had to be abandoned after pilot-testing in one district because what the students wrote indicated they could not treat the subject seriously. For example, teachers in one school reported derisive laughter and direct complaints, even refusals to write. Students at that school saw their

principal as an adversary, uncaring and unwilling to listen to recommendations. At a second school in the district, whose principal was especially effective, a homogeneous and complacent student body complained that they couldn't think of a problem worth writing about. Their awareness of the triviality of the problems they chose was reflected in their writing.

One potentially realistic audience specification which has elicited strong essays in some settings is the proposal that students write for readers of the school newspaper, student magazine, or yearbook. Yet in schools where there is no student press or where it is of poor quality, good writers may refuse to write or may write badly when asked to imagine they are writing for this audience.

Initial research in the composing processes of college writers (Flower, 1979) appears to suggest that specifying audiences is appropriate when the writing task is to communicate for business and social or political purposes. In the real world, such tasks have audiences whose special questions or needs influence the writer's choice of what to say and how to say it. In Flower's study, good writers felt more confident if they were given a particular audience when asked to write on an informative or persuasive topic. Poor writers paid little attention to audience specification or found that it made the task harder. In this research on the composing process of good and poor adult writers, Linda Flower asked subjects to write for or against on the issue of abortion for readers of the Catholic Weekly or the Children's Digest. In another study she had teachers and engineers describe their jobs for readers of Seventeen magazine. For self-expressive or literary tasks, specifying audience seems to make little difference, as if students recognize that a good story is a good story--no matter who the listener or reader. (Woodworth and Keech, 1980.)

In the Los Gatos Study\* (Keech), the writing task provided a particular audience and a clear rhetorical purpose for relating personal experiences. This task divided students into those who simply ignored the audience constraint, thus writing the kind of personal anecdote the test maker expected, and those who attempted to make a real transaction of the piece, giving themselves a much more difficult task. The versions of the prompt tested were:

---

\* See Chapter 2.

PROMPT 51a

VERSION I (implied audience)

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Write about this "first time" experience so that your reader can understand your feelings and why this memory has stayed with you.

PROMPT 51b

VERSION II (specified audience, bracketed here but not in student version)

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. (Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing.)

One student, composing aloud for the researcher, revealed that the audience constraints in the second topic meant, for her, that she had to give definite advice to a friend who was about to babysit for the first time. She tackled the task with sophistication, but in a thirty-minute test, her results were no match for the easy anecdotal essays of students who had version I, or who treated version II exactly like version I. The oral protocol of this student reveals complicated and mature strategies for writing to inform in a useful way, but the time provided was simply not adequate for her to finish the task she had set for herself. Other writing from the same student revealed she was perfectly capable of dashing off clever anecdotal writing, but the instructions in this case had seemed to her to call for something more.

- There are no simple answers to the question, "Should the writing prompt specify audience?" Strong reasons may be advanced for creating writing tasks that

specify audience and rhetorical purpose, especially when the tester wishes to call for persuasive or informative writing, but pilot-testing of such topics is essential to insure that students are helped rather than hindered when they take the presumed audience into account. The test-maker should consider two questions: Does specifying audience add information that is important to the writer in deciding how to respond to the task? Does the writing task really support the kind of writing a student might have to do or want to do, for an audience other than the teacher-tester who will actually be scoring the papers? If the answer to either of these questions is no, then the audience specification should be omitted.

## CONCLUSION

There is undoubtedly a great deal more useful information about what makes a good prompt residing in the heads of experienced essay test developers. Also, there are growing numbers of "homemade" writing tests being generated by the new widespread use of holistic writing assessments. But local test-makers complain of the lack of accessible information on how to make good prompts for writers, or how to prepare exacting instructions that direct in predictable ways the act of composing. To some extent, the lack of available "how to" guides is a result of the fact that the profession has yet to agree on what constitutes a "good" essay test. Writing good test topics remains an art, not a science, and criteria for judging a good topic may vary according to an evaluator's purpose.

It is becoming clear, both to full time and some-time evaluators, that more needs to be known about the effect of prompts on the writer being tested. Much existing folk wisdom about what makes a good prompt needs to be tested. Some popular practices need to be carefully qualified before being further promulgated. And, as with objective tests, makers of these direct measures need to understand the seriousness and complexity of their task; that is, the extent to which the prompts they give will influence the outcome of their assessments. They need to be alerted to unresolved problems in test design that concern all makers of direct measures of writing ability. Educators newly charged with designing writing tests that directly measure performance can be helped to gain this heightened awareness of the difficulty of their task. They can also perhaps become engaged in contributing new insights to test design, if they can become acquainted with some of the experience and insight of their predecessors. It is in the hope of calling for a fuller literature on the problems and best practices in test design for direct measures of writing ability that this tentative technical report is offered.

## References

- Braddock, R., R. Lloyd-Jones, and L. Schcer, Research in written composition, National Council of Teachers of English, Urbana, Illinois (1963).
- Camp, Gerald, Santa Clara County writing assessment, Bay Area Writing Project, University of California, Berkeley (1979).
- Caplan, Rebekah and Catharine Keech, Showing-writing a training program to help students be specific, Bay Area Writing Project, University of California, Berkeley (1980).
- Coffman, W.E., "On the reliability of ratings of essay examinations in English," Research in the Teaching of English, vol. 5(1), pp. 24-37 (1971).
- Cooper, Charles R. and Lee Odell, Evaluating writing, National Teachers of English, Urbana, Illinois (1977).
- Cooper, Charles and et al., Writing abilities of regularly-admitted freshmen at SUNY/Buffalo, Unpublished manuscript, San Diego (1979).
- Diederich, Paul, Measuring growth in English, National Council of Teachers of English, Urbana, Illinois (1974).
- Flower, Linda and John Hayes, "Problem-solving strategies and the writing process," College English, vol. 39(5), pp. 449-461 (December 1977).
- Flower, Linda, "Writer-based prose: a cognitive basis for problems in writing," College English, vol. 41(1), pp. 19-37 (September 1979).
- Gibson, Walker, Seeing and writing, Longmans, Green and Company, Inc., New York (1959).
- Gray, Stephanie and Catharine Keech, Writing from given information, Bay Area Writing Project, University of California, Berkeley (1980).
- Kirrie, Marjorie, "Prompt writing is not impromptu," National Writing Project Newsletter, vol. 1(3) (May 1979).
- Mullis, I., The primary trait system for scoring writing tasks, National Assessment of Educational Progress, Denver, Colorado (1974).
- Myers, Miles, A procedure for writing assessment and holistic marking, ERIC Clearinghouse on Reading and Communication Skills and the National Council of Teachers of English, Urbana, Illinois 61801 (1980).

Powell, David, What can I write about, National Council of Teachers of English, Urbana, Illinois (1981). 7,000 Topics for High School Students

Spandel, Vicki and Richard J. Stiggins, Direct measures of writing skill: issues and applications, Northwest Regional Educational Laboratory, Portland, Oregon 97204 (1980).

Stahl, Abraham, "The structure of children's compositions: developmental and ethnic differences," Research in the Teaching of English, vol. 2, pp. 156-164 (1977).

White, Edward, Comparison and contrast, California State Universities and Colleges, Office of the Chancellor (1973).

Woodworth, Patrick and Catharine Keech, The write occasion, Bay Area Writing Project, University of California, Berkeley (1980).

Jefferson County writing sample report" (1976).

The assessment of the writing performance of California high school seniors, California State Department of Education (1977).

The nature and measurement of competency in English, National Council of Teachers of English, Urbana, Illinois 61801 (1981).

## ACKNOWLEDGEMENTS

We are grateful to the number of professionals in testing agencies, the university professors and researchers in writing, and the Bay Area Writing Project teacher--consultants who provided written comments on an earlier version of this report. The Bay Area Writing Project teacher--consultant critic readers included Ruby Bernstein, Kate Blickhahn, Mary Frances Claggett, Jean Jensen, Irving Peckham and Bernard Tanner. The other professional critic readers included Evans Alloway, Gordon Brossell, Rex Brown, Charles Cooper, Kim Davis, Paul Diederich, Edmund Farrell, Sarah Freedman, Ann Humes, Ina Mullis, Bill Robinson, Joe Steele, and Edward White.

We are also grateful, to Marta Guzmán and Georgie Kelm who typed the several versions of the manuscript and who patiently tracked down elusive reference information. And finally, our thanks to Leo Ruth and Deborah Ruth for editorial work and suggestions.

## APPENDIX A

### SUGGESTIONS FOR WRITING ESSAY QUESTIONS

Gertrude Conlan  
Educational Testing Service  
Princeton, New Jersey

1. The question should be clear.

Students should not have to puzzle over the instructions. The topic is intended to test the ability to write the answer and not the ability to guess what the test maker intends. Besides, students have only a limited amount of time, time that should be spent writing and not analyzing unnecessarily.

2. The question should be as brief as clarity allows.

Restatement may sometimes be necessary to avoid misunderstanding. But, then, perhaps one should consider whether the restatement should be used without the original because the restatement does not need additional clarification.

3. The instructions should be definite.

Students should know what is required. For example, Discuss, citing specific examples from one novel. Or, Pay attention to the correct form of the business letter. Or, Be sure to use complete sentences.

4. Avoid questions requiring only a yes or no answer.

Example: Do you agree? Where does the student go from there?

5. Average students should be able to write average answers to the questions, and yet bright students should be able to show their brightness.

A good topic permits the ranking of all students according to ability.

6. The vocabulary used and the concepts expressed in the topic should not be too difficult for the ordinary student to understand immediately.

A difficult topic distinguishes only between the very bright and the rest of the population. Besides, difficult reading changes the test to a reading test.

7. The question should not call for clichés as answers.

A topic worn out by overuse produces worn-out responses. On the other hand, some good questions merely twist clichés. For example, What's right with television? In what ways are teenagers more conservative than the over-thirties?

8. The question itself should provide an organizing principle for the essay.

Example: Compare and contrast...  
Briefly describe...and then  
analyze...  
Discuss your answer to this question,  
giving the reasons for your answer  
and citing specific examples to  
support those reasons.

9. The question should not elicit responses which interfere with either the writer's or the reader's judgment.

Politics, racial issues, and other inflammatory topics are to be avoided. Also to be avoided are topics that are dishearteningly dull. For example, even a seemingly innocuous topic such as Who has had the greatest influence on your life? cannot be scored easily. If the candidate writes on the wrong political figure (from the reader's point of view), the score is either too high --

because the reader is making up for his or her own bias -- or too low because the reader has succumbed to that bias. On the other hand, the fifth essay on the greatness of the basketball coach is not scored on the same standard as the first. Readers are human; they do become bored.

10. The question writer should write out the answer expected and determine whether the question really calls for that answer. The question writer should also try to answer the question in the allotted time, just to see whether it is humanly possible to do so. The question should be revised in the light of any discoveries made.

1979

SOME CONSIDERATIONS IN JUDGING  
THE WORTH OF AN ESSAY TOPIC

Gertrude Conlan

1. Do the candidates seem interested in the topic?  
Do they all have something to say in response to it?  
Can they find specific examples to use?
2. Are the responses unduly emotional or patriotic or religious or in some other way difficult for readers to judge fairly?  
Can readers make judgments easily between essays at different levels of competence?  
Are all the essays alike -- mediocre?  
Does the topic encourage cliché-ridden or "canned" responses?  
Does the topic invite candidates to try to guess at what attitude the readers want them to take toward the topic?  
Does it leave writers trying to take all possible positions because they fear their own views may meet with disapproval?
3. Is the topic confusing in some way?  
Are the requirements of the topic ambiguous?  
Is the topic subject to several accurate interpretations?  
Will these create problems for scorers, by making comparison of quality difficult?  
Are the candidates misreading or misconstruing some word or phrase in the topic?  
Are candidates following the requirements of the topic?  
Is the mode of writing the candidates use in keeping with the mode the topic is meant to produce?
4. How will readers react to the topic?  
Are the responses just plain boring?  
Can judgments be made quickly on reasonable grounds?  
Are high scores being awarded for writing ability and not for extraneous matters like sophisticated literary allusions or unique examples?
5. How long are the essays?  
Has too much time been allowed or too little?
6. Do the essays lack a starting point?  
Do the essays lack organization?  
Is the topic at fault for failing to suggest a

method of organization?

Is the topic at fault for over-suggesting a method  
of organization -- no variety?

7. Would you be willing to read essays on this topic  
for several days in succession?

Dated 1979

215

APPENDIX B

PROMPT INDEX WITH SOURCES\*

PROMPT 1a

Think about the word trees. Write about what the word suggests to you. (Derived from topic 32 below, unpiloted example.)

PROMPT 1b

Think about the word trees. Write a story in which trees play an important part. This may be a real experience or an imaginary story. You may even write as if you were a tree. Be sure that trees play some part in your story. (Adapted by C. Keech, unpiloted example.)

PROMPT 1c

Think about the word trees. Write an essay in which you explain one of several possible ideas this word may suggest to you. You may write about a special kind of tree, a special use of trees, people's attitudes toward trees, or something else that interests you about trees. Focus on one idea--do not attempt to list all the things trees can mean. (Adapted by C. Keech, unpiloted example.)

PROMPT 2a

g Write about December 25, 1973.

PROMPT 2b

Write about an experience you had during a family holiday which was particularly memorable. It may be happy or sad, but should be a time when you had strong feelings. (By C. Keech.)

PROMPT 3

\*Note: The prompts in the following list have all been used in this report to illustrate certain principles for developing prompts. Thus, they should not be used without consideration of the assessment context described in the report. In addition, some of them are considered to be faulty in the form presented, and these are marked by the null sign @.

Write about a place that is special to you. Help your reader understand its peculiar interest or significance. The place you choose may be of any kind--for example, a personal haunt, a historical spot, or an institution or a place you remember vividly from early childhood. (Variation of prompt used in the Sir Francis Drake High School, San Anselmo, California school-wide assessment, 1968.)

PROMPT 4

Write about an object you are especially attached to, something that has deep personal meaning for you, something that has become a part of your life. You might want to consider the way you discovered it, the way it came into your life, the way it has taken on meaning through time. (Variation of prompt used in the Sir Francis Drake High School, San Anselmo, California, school-wide assessment, 1972.)

PROMPT 5

Tell about something you enjoyed doing as a child. (Caplan and Keech, p. 48, used in research study.)

PROMPT 6

Tell about an experience from which you learned a lesson. (Caplan and Keech, p. 47, used in research study.)

PROMPT 7

Think of someone you admire or resent. (Not your parents.) Write about this person and include incidents that led you to admire or resent this person, so the reader can understand why you feel the way you do. (Las Lomas High School, Walnut Creek, California, 1977 school-wide assessment. Edited version.)

PROMPT 8

Describe a change that has happened in you recently or an event that happened a long time ago but still affects you. Tell of the events that surrounded it and caused the change in you so your reader can understand what is now different about you and why. (Las Lomas High School, Walnut Creek, California, 1978 school-wide assessment. Edited version.)

PROMPT 9

Each of us assumes a personality to satisfy each group into which we enter; i.e., we do not act at home precisely as we act with the gang, in the classroom, or before an employer. (College Board Examination released topic--adopted version.)

PROMPT 10

- 0 Write about your favorite place. (A reduction of Topic 3.)

PROMPT 11

- 0 Write about your favorite object. (A reduction of Topic 4.)

PROMPT 12a

Think of an occasion when you experienced something for the first time. It may be something you later came to take for granted, but for some reason, the first time was memorable. Write about your experience so that your reader understands your feelings at the time and why it stays in your memory. (C. Keech, one of two versions used in Los Gatos High School, California experimental study, 1930.)

PROMPT 12b

Think of an occasion when you experienced something for the first time.

CUES:

There is a book called The First Time, in which famous people describe their first kiss or first love affair. But there are many other important First Times in our lives, from the first time we tied our own shoes, or went to school or camp, to our first time in deep water, or first big game or first speech in front of a group. Some of these occasions are first and only times. Others are the first of many times, but for some reason the first time was memorable.

INSTRUCTIONS:

Choose one of these occasions that stands out in memory and write about it so that your reader can

understand your experience and why it was memorable. (C. Keech, unpiloted example.)

PROMPT 12c

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing. (C. Keech, one of two versions used in Los Gatos High School, California experimental Study, 1980.)

PROMPT 13

Compare two stores you know which sell similar items. Consider such things as prices, merchandise, service, atmosphere, location, or other special features. Which of the two stores do you prefer? Explain why. (Gray and Keech, p. 26, used in research study.)

PROMPT 14.

Compare two actual holidays or special events you celebrate in your family (such as Thanksgiving, Passover, Christmas, the Fourth of July, your birthday, etc.) Consider such things as food, decorations, activities, behaviors, atmosphere, or other special features. Which of the two do you prefer? Explain why. (Gray and Keech, p. 26, used in research study.)

PROMPT 15a

You have met a man from outer space who has landed on earth near your school playground. He can understand English, but he does not know anything about school here on earth. Describe your school for him. Tell him about your school building. Tell how the school building looks on the outside and the inside. Tell him about your teacher and your classmates. (italics added) (California State Board of Education Handbook on Assessment, 1979.)

PROMPT 15b

Imagine a friend of yours is about to move to town. Your friend's name is Pat. Pat will be going to your school. Below, write a letter to Pat, describing one or two things you think a new student should know. You might tell about what you do at lunch time, or describe the games you play. You might describe your favorite part of the classroom or what your teacher allows you to do in class activities. (C. Keech variation of Topic 15a, unpiloted.)

PROMPT 16a

Write about your favorite place. Name the place and tell where it is. Tell how often you go there. Why is this place special to you? How do you feel when you are there? What are some sights, sounds, colors and smells in your favorite place? (San Mateo City Elementary School District, California, Courtesy of Joanne Graham.)

PROMPT 16b

Everyone has a special place where they like to go. Write about your favorite place to go. Describe the place so your reader can imagine what it is really like. Tell about what you do there, and how you feel when you are there. Help your reader see why this place is special. (Revision of 16a.)

PROMPT 17

Write a story that begins: 'One day I found a magic machine.' (Middle school teacher, American School of the Hague, 1975, unpiloted in assessment.)

PROMPT 18

This morning on my way to school, I saw a big black hole. (C. Keech, classroom assignment, unpiloted in assessment.)

PROMPT 19

The last thing I remember was the ship in flames, people screaming and running for lifeboats. Then I blacked out. When I came to, I was... (C. Keech, classroom assignment.)

PROMPT 20a

Dear Wish Fairy,  
I wish I had . . .

INSTRUCTIONS:

The teacher gives the following instructions orally:

"Pretend I am the Wish Fairy who can make your wishes come true. I want you to write me a letter to tell me what you wish for. Think of several things you wish for. (pause) Now then, choose the one thing you wish for the very most.

Tell me what it is by finishing the sentence, "I wish I had..." What do you wish you had? After you write down its name, then go on writing and tell me all about it. What is it like? Write down everything you think and feel about your wish. (Thanks to an unknown elementary school teacher who presented it at a California Association of Teachers of English Conference. Unpiloted in assessment.)

PROMPT 20b

"I wish I were..."  
"I wish I could..."  
"I wish I knew..."  
(Variations in common use.)

PROMPT 21

Write about an event you wish you had witnessed or could witness. The event can be real or imagined; the time of the event can be past, present, or future. Make it clear why the event is significant to you. (Variation of prompt used in Sir Francis Drake High School, San Anselmo, California annual assessment, 1973.)

PROMPT 22

If you had to choose to be something other than a human being, what plant or animal or other form would you choose? In your writing, give your reader some idea of what you think it would be like to be that form, and of why you chose it. (Variation of prompt used in Sir Francis Drake High School, San Anselmo, California annual assessment, 1974.)

PROMPT 23

If you could change places with someone else, who would it be? The person you write about can be living, dead, drawn from past or present, from books, films, etc., or from your own imagination. In your writing give your reader some idea of what it would be like to be that person, and of why that life appeals to you. (Variation of prompt used in Sir Francis Drake High School, San Anselmo, California annual assessment, 1975.)

PROMPT 24

What would you do if you had a million dollars?  
(Topic in common use.)

PROMPT 25

While visiting in your grandmother's house you find a large trunk. Propped on a chair beside the trunk is a card; on it in your grandmother's handwriting are the words: "To my only grandchild: Inside this trunk you will find what you have always wanted." Describe the gift and how you will use it. (Santa Clara County Field Test of Prompts, 1981, p. 6.)

PROMPT 26

Imagine that you are trapped in one of the following situations and you cannot escape for several hours. Tell what you do to escape or to pass the time until you are rescued:

- A. an elevator
- B. an abandoned mine
- C. a locked room
- D. a life-boat

(Santa Clara County Field Test of Prompts, 1981, p. 5)

PROMPT 27

Imagine that you have just landed on the moon. Write about what happens after you climb out of your space ship and set foot on the surface of the moon. (Suggested by Michael Scriven, piloted.)

PROMPT 28

Pretend you write the advice column in our school newspaper. You must answer the following letter:

Dear Happy Helper:

I have a little sister who always hits me when I have a toy or a cookie she wants. My grandmother says I can't hit her back. She says I should share, but I get tired of sharing. What do you think I should do? (C. Keech, unpiloted in assessment.)

Signed,

In Trouble

PROMPT 29

- Ø Explain your agreement or disagreement with the following statement: The history of women in our country is the history of oppression. (Source lost.)

PROMPT 30a

- Ø "Men are superior to women." Agree or disagree with this statement, giving reasons for your opinion.

PROMPT 30b

- Ø "Women are superior to men." Agree or disagree with the statement, giving reasons for your opinion.

PROMPT 30c

- Ø "Women are better off than men in today's society." Agree or disagree with the statement, giving reasons for your opinion.

PROMPT 30d

Explain what is wrong with the following argument: "It's pretty clear that on the whole, women are superior to men. In the physical dimension, the best evidence for this is simply that women live longer, and by a substantial number of years. Other evidence is that women are virtually immune to several diseases that are fatal for many men; and on the mental side, there is the simple fact

that girls, on the average, do better at school than boys." (Michael Scriven for Carnegie Evaluation of Bay Area Writing Project.)

PROMPT 31a

At the place where you work, a woman has just quit her job, leaving vacant the company's only executive position ever held by a female. The Board of directors has stated their preference that a woman replace her in order to fulfill an Affirmative Action quota. As a member of the Hiring Committee, it is your job to help choose a successor to the post.

The only woman who has applied for the job seems competent and meets the written qualifications for the job but she is clearly less qualified than both of the men she is competing with. Members of the Committee disagree about what should be done: some say hiring a woman is absolutely necessary for breaking down employment discrimination; others say hiring a less qualified person would be foolish as well as unfair to those working under the new executive.

To have a full hearing of all views on this critical issue, the Hiring Committee has asked each member to prepare a carefully written statement to be distributed in advance of a meeting to discuss the issue. Write a statement which represents your position in the matter, making it as logical and persuasive as possible. Your writing task is to persuade the Committee to adopt your own view and to vote on the job candidates in accordance with your view. (Charles Cooper, Writing Abilities of Regularly-Admitted Freshment at SUNY/Buffalo, 1979.)

PROMPT 31b

- Q Argue pro or con affirmative action policies in admissions or employment. (C. Keech variation of 31a.)

PROMPT 32

- B Consider the words trees and plastic. Write a paragraph in which you discuss some aspect(s) of what the words suggest to you. (Marjorie Kirrie, Portland University, Oregon.)

PROMPT 33

Even those of us who may conform outwardly have our private, personal forms of rebellion. Some rebel against society's conventions, some against family values, and others against the attitudes of peers. We might dress in unconventional ways, reject our family's traditions, or disagree with friends' political beliefs. Some of us may simply rebel against popular fads. Many of us are not interested in jogging five miles a day, eating natural foods, or wearing designer jeans. (Jo Keroes, San Francisco State University.)

PROMPT 34

"Man is a creature of habit."

Do our habits make us more stable, more efficient, or just duller? Discuss. Support your arguments with one or two specific illustrations from your reading or experience. (Adapted from released College Entrance Examination Board Topics, 1963-1966.)

PROMPT 35

"Opposition is indispensable. A person will often learn more from the people who disagree with him or her than from those who agree."

Do you agree with this observation? Have you profited more from people whose views were similar to yours or from people whose views opposed your own? Discuss, using one or two specific examples to support your position. (Adapted from released College Entrance Examination Board Topics, 1963-1966.)

PROMPT 36

Not all inventions have been good for humanity. Name one invention we would be better off without, and make it clear why. You may do this writing as an essay, journal, letter, story, or other form. (Variation of prompt used in Sir Francis Drake High School, San Anselmo, California annual assessment, 1976.)

PROMPT 37

Imagine that a small group of people will be sent to colonize a new planet. Food, clothing, shelter and transportation have been provided for. You are among those asked to select a few additional things to be sent along in the limited space available in the ship. What one item would you recommend, and why? You may write your recommendation in the form of a story, a dialogue, a letter, a speech, an essay or other form. (Sir Francis Drake High School, San Anselmo, California annual assessment.)

PROMPT 38

Describe your favorite \_\_\_\_\_ (person, food, object) so that your reader will understand why it is your favorite. (Commonly used topic.)

PROMPT 39

My Home: Write a description of your home, its rooms and their contents in such a way that someone who has not visited it can form an idea of the way it looks. (Abraham Stahl, Research in the Teaching of English, 1977, suggested by Leo Ruth.)

PROMPT 40

"Many observers of our society claim that modern man, immersed in materialism, is 'owned by his objects.' Yet many of us have objects that we treasure not just for their material value but for a variety of other reasons. Assignment: Describe one or more objects which are important to you. Explain what values they represent and comment on those values." (Edward M. White, The California State University and Colleges Freshman English Equivalency Examination, 1975--sample essay question.)

PROMPT 41

"Find a church in your neighborhood that seems to you an interesting building. Describe what you see on its facade in detail and in such a style that your reader sees it as interesting too." (Walker Gibson, Seeing and Writing, 1959, classroom assignment, unpiloted in assessment.)

PROMPT 42a

"Describe what you see." (Accompanying a photo and diagram of an automobile accident.) (Based on a 1976 California State Assessment topic.)

Imagine that a small group of people will be sent to colonize a new planet. Food, clothing, shelter and transportation have been provided for. You are among those asked to select a few additional things to be sent along in the limited space available in the ship. What one item would you recommend, and why? You may write your recommendation in the form of a story, a dialogue, a letter, a speech, an essay or other form. (Sir Francis Drake High School, San Anselmo, California annual assessment.)

PROMPT 38

Describe your favorite \_\_\_\_\_ (person, food, object) so that your reader will understand why it is your favorite. (Commonly used topic.)

PROMPT 39

My Home: Write a description of your home, its rooms and their contents in such a way that someone who has not visited it can form an idea of the way it looks. (Abraham Stahl, Research in the Teaching of English, 1977, suggested by Leo Ruth.)

PROMPT 40

"Many observers of our society claim that modern man, immersed in materialism, is 'owned by his objects.' Yet many of us have objects that we treasure not just for their material value but for a variety of other reasons. Assignment: Describe one or more objects which are important to you. Explain what values they represent and comment on those values." (Edward M. White, The California State University and Colleges Freshman English Equivalency Examination, 1975--sample essay question.)

PROMPT 41

"Find a church in your neighborhood that seems to you an interesting building. Describe what you see on its facade in detail and in such a style that your reader sees it as interesting too." (Walker Gibson, Seeing and Writing, 1959, classroom assignment, unpiloted in assessment.)

PROMPT 42a

"Describe what you see." (Accompanying a photo and diagram of an automobile accident.) (Based on a 1976 California State Assessment topic.)

PROMPT 42b

Study the accident in the photo, using the diagram to help you gain a clear understanding of the situation. Describe the accident, reporting details accurately so that a jury or insurance company can make a fair decision about who is to blame. (Variation of Topic 42a. Idea from Beth Breneman.)

PROMPT 43

The following advertisement appeared in the local paper:

SUMMER EMPLOYMENT

Earn & Learn

On-the-job training for future positions in:

Medical Services Aide at Highland Hospital  
Ticket taker at local theatre  
Grocery clerk at local supermarket  
Plumbing Assistant for plumbing company  
Assistant to the manager of a circus

Earn \$3.50 or more per hour while you learn a valuable skill.

Send letter of application to:  
Ms. Laura Jones, Opportunity Training Center  
2212 Fruitvale Ave., Oakland, California 94231

Pretend that you are applying for one of the jobs listed. Write a letter explaining why you should be hired. You should give information about yourself. The information may be actual facts, made-up information, or a combination of fact and fiction. The letter should help you get the job. (Oakland School District Assessment, Oakland, California.)

PROMPT 44

Describe your school. Tell how it looks from the outside. Tell how your classroom looks. You may describe what happens in the hallways at different times of day, or you may describe the lunchroom or playground. If you like, you may describe the people in your school. (Source unknown.)

Ø Look carefully at the picture.

These kids are having fun jumping on the overturned boat. Imagine you are one of the children in the picture. Or if you wish, imagine that you are someone standing nearby watching the children. Tell what is going on as he or she would tell it. Write as if you were telling this to a good friend, in a way that expresses strong feelings. Help your friend feel the experience too. Space is provided on the next three pages. (National Assessment of Educational Progress, Report No. 05-W-02, Expressive Writing, 1976.)

PROMPT 46a

Look at the picture above. You can react to it in any way you wish by writing about it in one of the following ways: as a story (What have the characters been doing? What could happen next?), as a description, or as an opinion. (Jefferson County, Colorado, Writing Sample Report, 1976.)

PROMPT 46b

Look at the picture above. Let it give you ideas or create a mood. Then use the ideas or the mood to write a story, a memory of something that happened to you, or an essay or argument based on some strong feeling inspired by the picture. (C. Keech revision of Topic 46a.)

PROMPT 47

Ø "No man is an island, entire of himself. Every man is a piece of the Continent, a part of the main..."  
John Donne

This topic should lead you to think about the possible meaning of the quotation. You may build upon the idea by supporting it, or you may contradict the topic with an opposing point of view. Attempt to build a logical, reasonable opinion paper. Include an introductory and concluding paragraph and two or three paragraphs in the body of the paper. The supporting paragraphs may be developed by comparison/contrast, definition, illustration, or any other form which will produce a unified essay on this topic. (Jefferson County, Colorado, Writing Sample Report, 1976.)

PROMPT 48a

9 "Write a letter to your principal. Explain to him or her what you would like to see changed in your school." (C. Keech, example of problem)

PROMPT 48b

"There are probably several things you would like to see changed in your school. CHOOSE ONE OF THESE and write a letter to your principal, describing or explaining the problem and suggesting what he or she might do to change the situation." (Variation of NAEP "letter to the principal" Writing Report, 05-W-03, 1977 by C. Keech.)

PROMPT 48c

There is probably at least one thing about your school that you would like to see changed. Write a letter to your principal and tell him about the problem. Be sure to describe the situation as it is now and tell why it bothers you, as well as how you would like to change it. Variation of NAEP "letter to the principal", Writing Report, 05-W-03, 1977 by C. Keech.)

PROMPT 49

The following article appeared in the local newspaper:

KDOL Begins Survey:

Why do we remember certain people from our past and forget others? Television station KDOL is preparing a special TV show on personalities people remember from their elementary school years--a classmate, a member of the school staff, a family member, or a neighborhood friend. Send your description to:

Program Producer  
KDOL  
Oakland, CA 94606

Respond to the article by choosing someone who made an unusually strong impression on you. Write about that person, explaining why you remember him or her. Do your best work. (Miles Myers, Oakland Schools Assessment.)

PROMPT 50

"Write a friendly letter to a pen pal in another country, telling him or her about your home or school." (Commonly used topic.)

PROMPT 51a

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Write about this "first time" experience so that your reader can understand your feelings and why this memory has stayed with you. (C. Keech.)

PROMPT 51b

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing. (C. Keech).

Chapter 2

EFFECTS OF VARIATION IN A WRITING TEST PROMPT  
UPON HOLISTIC SCORE AND OTHER FACTORS

Catharine Keech, Don Leu, Sandra Murphy,  
Charles Kinzer

Part I: Effects of Two Versions of a Writing  
Prompt Upon Holistic Score and Writing  
Processes

Don Leu, Catharine Keech, Sandra Murphy,  
and Charles Kinzer

Part II: Comparison and Analysis of Rater Responses  
to the Anchor Papers in the Writing Prompt  
Variation Study

Catharine Keech and Mary Ellen McNelly

## Chapter 2: Part I

### EFFECTS OF TWO VERSIONS OF A WRITING PROMPT UPON HOLISTIC SCORE AND WRITING PROCESSES\*

Don Leu, Catharine Keech, Sandra Murphy,  
and Charles Kinzer

#### OVERVIEW OF THE STUDY

To investigate effects of variation in the specification of audience on a timed writing test, 114 high school students, grades 10-12, from four classes of one teacher, were randomly assigned one of two versions of a thirty-minute test asking them to write about a particular personal experience: an occasion when the writer experienced something for the first time. The subject and the time limit are typical of writing tests used in assessments of writing ability by many local school districts. The essays were ranked ordered by quality following holistic rating procedures.

Data about the subjects were collected to provide insight into additional factors other than the task variations which might also affect the quality of performance as measured by the holistic score. Estimates of verbal skill and attitude were obtained by having the teacher rate subjects individually on the basis of classroom behavior and performance. Data on sex and academic proficiency level were also recorded. A proficiency level was assigned on the basis of the subjects' grade level and academic track placement.

Finally, the day after writing in response to one of the two versions of the "first time" prompt, subjects in the study completed a questionnaire designed to elicit information about their attitudes toward the prompt, their recollections of the writing process during the timed writing test, and their descriptions of their other writing behaviors and attitudes, both in and out of school.

Associations between holistic score, version of prompt, teacher rating of verbal skill, teacher rating of attitude, sex of the writer, proficiency, and responses to questionnaire items were examined. Two questions were of particular interest:

---

\*This report is based upon research supported by the National Institute of Education under Grant No. NIE G-80-0034 to the Bay Area Writing Project, University of California, Berkeley, Leo Ruth, Project Director.

- 1) Which variables are associated with high holistic scores?
- 2) Which variables are associated with differences in the prompt?

The data yielded interesting information in response to both questions, suggesting differences in prompt affected specific aspects of the writing process although the mean quality of performance as measured by holistic scores was similar for both versions. Unexpected patterns of response on the questionnaire, for example, indicated differences in the planning process with an increase in planning time associated with the minimal audience specification.

Although one task constraint, degree of audience specification, was manipulated experimentally, the study as a whole should be viewed as descriptive and explanatory rather than as an exactingly controlled experiment. Viewed in this light the study offers a number of provocative leads for needed investigations.

## PROCEDURES OF THE STUDY

### Subjects

The subjects of the study were 114 students at Los Gatos High School, located in Los Gatos, an upper-middle class, suburban community in Northern California. The subjects were students in four classes taught by one teacher whose program differs somewhat from the traditional English program generally followed at the school. The traditional program emphasizes the teaching of grammar and instruction in basic skills with a concentration on the forms of exposition in upper grade level courses. The teacher in this study, a teacher-consultant with the Bay Area Writing Project, Mary Ellen McNelly, emphasizes writing from personal experience, providing more opportunities for expressive writing than might be found in other classes at the school.

The students in the school are tracked, according to ability level, into accelerated and non-accelerated English programs. Tracking decisions are based largely on two factors: teacher recommendations and performance on a test constructed by the English department. The departmental test consists of both an objective examination and a writing sample. The objective examination covers usage, mechanics, vocabulary, and spelling. The writing sample is evaluated on the basis of both form and content.

The 114 students who make up the population of this study came from one section of advanced composition juniors and seniors ( $N = 22$ ), two sections of college-track English literature juniors ( $N = 65$ ), and one section of general English, average-ability sophomores ( $N = 27$ ). The 87 juniors and seniors in advanced composition and English literature were in the accelerated program at the school. The 27 sophomores in general English were in the non-accelerated program. There were 66 female subjects and 48 males. The number and percentage of students in each category can be seen in Table 1.\*

### Determination of Verbal Skill and Attitude Ratings

The teacher provided subjective impression ratings of students' verbal skills and attitudes toward achieving in school, two factors which might be expected to be associated with performance on a writing test. At the end of the school year, when the testing took place

---

\*All tables and figures appear at the end of this report.

(May 1981), the teacher could be expected to have formed a clear impression of her students' capabilities, especially since her particular teaching style was informal, personal and interactive. A teacher with 20 years of experience, she could also be expected to have established intuitive "norms" or expectations for performance at a given age level. The teacher was asked to draw on her total impression of a student gained from what she had observed of his or her ability to comprehend literature and to express himself or herself orally and in written English on a variety of occasions, both formal and informal. She was then asked to assign each student a rating from 1-4 on a factor called general verbal ability: 1 = well below grade level; 2 = grade level or below; 3 = grade level or above; and 4 = well above grade level. She was also asked to assign each student to one of four categories based on her evaluation of the student's attitude toward achieving in school: 1 = extremely negative, 2 = mildly negative, 3 = mildly positive, 4 = extremely positive. A "positive attitude" would indicate that the student appeared motivated to perform well in school and seemed to be working to his or her best ability.

#### Timed Writing Test Versions

This writing test occasion required students to compose a response to one of the following two versions\* of a personal narrative prompt within thirty minutes:

#### PROMPT VERSION A:

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Write about this "first time" experience so that your reader can understand your feelings and why this memory has stayed with you.

You may write an essay, a short story, a

\*The variable sets of instruction in the two prompts are presented here enclosed in boxes for purposes of comparison in this report. The students, however, received the test instrument in the format appearing in Appendix A, with only one version of the prompt appearing in the writing test.

letter, or a journal entry as a way of re-telling your experience.

PROMPT VERSION B:

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing.

You may write an essay, a short story, a letter, or a journal entry as a way of re-telling your experience.

The personal narrative topic, "Tell about a first time experience", was chosen as being typical of prompts used in high school writing assessments scored holistically. Such topics are often used because they assure that almost every student will find material for writing in personal experience and be able to begin composing quickly. These prompts aim to elicit samples of general fluency in written English, drawing on material students may have already told orally. Because this prompt, especially Version A, does not require special writing skills such as argument or exposition, it can be used across several grade or ability levels and in different courses.

The major difference between the two versions of the prompt is the more explicit specification of the audience and rhetorical purpose in Version B. Both versions however, ultimately are directed to an implicit audience of test readers because the writing occurs in the context of a testing situation. In Version B the instructions specify an imaginary audience, "someone" who has shared or will share the first time experience the student chooses to write about. In Version A specification of the audience is less restricted and refers simply to "your reader."

This study of effects of variation in prompt

wording builds on a study reported in The Write Occasion (Woodworth and Keech, 1980) which set out to test a common assumption by test-makers that specifying the audience and rhetorical purpose enhances the quality of student performance. That study failed to find a difference in mean holistic scores between the two versions, but hypothesized several explanations for this result. The present study probes further for differences that may not be reflected in holistic scores, through the administration of a post-test questionnaire. One important reason for continuing to examine variations in the wording of prompts even when they do not appear to significantly affect holistic scores is to determine how student writers construe the writing task. The research goal is to identify possible constraints affecting the process of interpretation of task demands by student-writers and evaluators of the papers elicited. This knowledge will contribute to the development of a model of the whole writing assessment episode.

#### Administration of the Test and the Holistic Rating

The two versions of the prompt were distributed randomly by the regular teacher in the four classes (in May 1981) so that equal numbers of students in each group wrote on each version. The teacher distributed the tests and read through the instructions (See Appendix A for complete instructions to students) with students, pausing for students to write their individual code numbers. Test conditions were consistent for all four classes. Ten students were taken out of the classes to do the writing test under special conditions requiring oral composing, under the supervision of the investigator, Catharine Keech.\*

Each paper received three sets of holistic scores.\* First, the papers were randomly mixed across all classes and grade levels, then distributed to peer raters, who had received special training and who were the first holistic raters of the essays. A second set of scores was provided later when the researcher Catharine Keech led a group of student-teachers at the University of California, Berkeley in rescoring the

---

\*The ten oral composing protocols thus elicited have not been analyzed, but the ten papers produced are included in the reporting of all results below, having been scored as part of the larger sample of papers.

\*The three rating sessions are described in Part II of the report where results of the training sessions are discussed and comparisons of the three sets of scores are reported.

papers, using training techniques similar to those she had used with the original student raters. A third set of scores was provided shortly afterwards by four "expert" holistic raters who were teacher/consultants for the Bay Area Writing Project. This third set of scores, assigned by the expert raters in a blind ranking, provides the scores for the analysis reported below.

In general, the training of the expert holistic raters followed conventional procedures. The four experts ranked anchor papers selected by the researcher Catharine Keech, acting as trainer. Scores were assigned by each rater on a scale of 1-4, with 1 assigned to the poorest papers and 4 to the best papers. The anchor papers showing the quality of writing at each scoring level are included in Appendix B. Since each paper was scored by two raters, the total score for any paper could range from a minimum of 2 to a maximum of 8.

The expert raters quickly reached consensus on the criteria to be applied in rating the entire sample. They achieved a significant level of reliability: the Pearson product moment correlation between the first and the second scorer was  $r = .71$  ( $p < .01$ ). Discrepancies (two scores given to the same paper that are more than one point apart) were resolved in the usual way in a third reading by the researcher acting as the table leader.

#### Administration of the Questionnaire

Dittoed copies of a typed questionnaire consisting of 25 items were distributed to the students during their regularly scheduled class periods on the day after the writing test. Complete questionnaire data were collected on 99 of the 114 students.

The questionnaire was divided into three sections. In the first section, students were asked to read three versions of the prompt. Two of the versions had been assigned for the writing test (Prompt Versions A and B above). The third version was new to the entire group:

#### PROMPT VERSION C

There is a popular book called My First Time, edited by Bennett Cerf, which is a collection of pieces by famous people who are describing experiences they remember when they were doing something for the first time. They tell about first kisses, first time on the stage, first crime, first time they saw

someone famous they would later work with, first political campaign, first time they were afraid, etc. Some of their experiences are unusual, some are common but well told."

I would like to collect a book of First Time experiences of high school students. I have asked a number of teachers to allow their students to write about their First Times during a 30 minute writing test, from which I will select promising pieces. Later I will contact the authors of the selected pieces and give them an opportunity to revise and polish before publication. Chosen pieces will be copyrighted and the authors will receive \$50 for their contribution. No piece will be used without the author's permission.

You may write an essay, a short story, a letter, or a journal entry as a way of re-telling your experience.

Students were asked to identify which version they actually had been given, and also which version they might have preferred to write on if they had been given these three choices. Space was provided for comments regarding preference.

In the second and third sections of the questionnaire, students responded to statements (e.g. "I consider myself a good writer.") on a four-point Likert scale: strongly agree, agree, disagree, strongly disagree. A copy of the complete questionnaire may be found in Appendix B.

#### Choice of Statistical Tests

In reporting agreement between readers during holistic scoring, Pearson's product moment correlation ( $r$ ) is used as a measure of association comparing scores given by the first and second readers on every essay, following common practice in holistic assessment.

With the exception of that special use of Pearson's  $r$ , the statistical measures used in this report are non-parametric, specifically tau and chi square ( $X^2$ ). Non-parametric measures are appropriate in research where the number of cases changes across comparisons and where those comparisons were not planned as a part of the design of the study, but are more in the nature of tentative seekings of possible

relationships. Because the comparisons are unplanned, in this sense, rather than part of a tightly controlled hypothesis testing, it is important to use the most conservative possible estimates of association of variables. Non-parametric statistics serve this purpose, making it more difficult to find statistical significance. Tau is a measure of association, like  $r$ , expressing a relationship between an independent and a dependent variable in terms of a range of values from .00 to 1.00, with lower values representing a weak degree of association and higher values representing a strong association. It is appropriate when both variables provide a linear set of values from lower to higher numbers.

Chi Square ( $X^2$ ) is not a measure of association in the same sense; rather it is a simple yes-no test for the non-independence of two discrete, categorical independent variables, each having only two values, + and -. The value of  $X^2$  must be greater when the number of cases is smaller, in order to demonstrate the non-independence (or association) of two variables as statistically significant. In reporting a relationship between two such variables, we give the  $X^2$  value in parentheses, together with the  $N$  of cases, to demonstrate statistical significance at the  $p = .05$  level. (These numbers do not represent degrees of association, as do the numbers reported as values for tau)

Responses to the questionnaire items can be most understandably reported in relation to other variables if the Likert four-point scale is reduced to a two-part categorical variable, with positive responses to an item set against negative responses. Since many of the variables being related to the questionnaire items are themselves categorical, chi square ( $X^2$ ) is the appropriate test to determine whether student attitudes as revealed in the questionnaire are in any way associated with other independent variables such as sex, proficiency (determined by placement in classes), mode of writing (collapsed into two categories, narrative and non-narrative), etc. It also appropriately tests relationships between items on the questionnaire.

## GENERAL RESULTS

The following sections provide descriptive data about the characteristics of the population tested and the relationships of these data to holistic scores. Interaction between variables and their relation to holistic scores are also reported here. And finally, the descriptive analysis of the questionnaire data is provided.

### Verbal Skill, Attitude, and Holistic Score

The population distribution for verbal skill rating, attitude rating, and performance on the timed writing test as measured by the holistic scoring are presented in Tables 2, 3, and 4. The distribution of ratings is positively skewed on all three variables, indicating that most of the students possess well-developed skills and favorable attitudes toward their class and work. This was expected since three of the four class sections are accelerated courses for highly able and ambitious students.

Significant associations were found between all three variables. The correlation between the teacher's estimate of verbal skill and attitude was significant and positive ( $Tau_b = .583, p < .05$ ). Significant positive correlations were also found between verbal skill and performance on the timed writing test as measured by holistic score ( $Tau_b = .33, p < .05$ ) and between attitude and holistic score ( $Tau_b = .35, p < .05$ ). An increase in holistic score, among these students, is associated with an increase in attitude rating and an increase in verbal skill rating; students with high holistic scores tend also to have high attitude and verbal skill ratings. Moreover, the association of attitude rating with holistic score is slightly stronger than the association of verbal skill and holistic score.

### Sex

No significant relationships were found between sex and holistic score, verbal skill rating, or attitude rating. Both males and females performed at the same level of proficiency on the timed writing test. Both sexes also had similar verbal skill ratings and attitude ratings.

### Proficiency Level

In this investigation, proficiency level was defined by each student's course placement; that is,

the juniors and seniors in the advanced composition class and the juniors in the two sections of English literature were in the accelerated track and were, therefore, considered to be relatively proficient writers. The sophomores in the general English class were in the non-accelerated college-prep track, and they were also a year behind the junior sections in their academic program; they can, therefore, be classified as less proficient writers. The validity of this classification is supported by a comparison of the median rankings of the two groups' total holistic scores using a Mann-Whitney U Test. Juniors and Seniors in the accelerated track performed significantly better on the timed writing test than sophomore students who were in the non-accelerated track ( $U = 520.00$ ,  $z = 4.56$ ,  $p < .05$ ). The mean rank of the proficient students was 65.02 compared to the mean rank of the less proficient students which was 33.26. The means and standard deviations of proficient and less proficient writers' total holistic scores can be seen in Table 5.

There was no relationship between the sex of the student and proficiency level\* ( $\chi^2 = 0.00$ ,  $p > .05$ ), meaning that proportions of males and females were similar in both proficiency groups. There were, however, significant relationships between proficiency and verbal skill rating ( $\chi^2 = 8.46$ ,  $p < .05$ ) and proficiency and attitude rating ( $\chi^2 = 19.98$ ,  $p < .05$ ).

Of proficient writers 70.6% (N=60) were considered above grade level in verbal skill compared to only 37.0% (N=10) of the less proficient writers. Of proficient writers 81.2% (N=69) were considered by their teacher to have a positive attitude toward achieving in school compared to only 33.3% (N=9) of the less proficient writers. Proficiency, then, seems to be importantly related to verbal skill and attitude in addition to total holistic score.

#### Effect of Prompt on Score and Selection of Mode

The two versions of the prompt were distributed equally among the population tested. Though the prompts were assigned randomly, each of the two groups thus constituted proved to be comparable in their distribution of sex, verbal skill and attitude of the students. Likewise, scores on the two prompts were comparable. There was no association evident between the

---

\*These analyses were conducted after collapsing the four verbal skill categories into two (above grade level and below grade level) and after similarly collapsing the four attitude categories into two (negative and positive).

assigned prompt and the holistic performance score among the 114 students who completed the writing test: the mean ranks of total holistic scores were not significantly different for the two versions ( $X^2_{\text{rank A}} = 54.97$ ,  $X^2_{\text{rank B}} = 60.03$ ,  $p > .05$ ). The mean score for prompt A was 5.09, the mean for prompt B was 6.28.

An examination of the students' papers showed that different genre were used in responding to the two versions of the prompt (e.g., narrative, letter, journal, essay). An analysis was then performed to find whether the prompt itself had any effect on selection of the mode of writing (see Table 6).

The narrative mode was the dominant choice of writers on both tasks. Even so, some writers made other choices. Although 76.3% of students wrote a narrative, 18.4% wrote a letter, 4.4% wrote a journal entry, and 0.9% wrote an essay in responding to the prompts. In all, a total of 27 students (23.7%) chose alternate, or non-narrative modes of writing.

The effects of the prompts are clearly associated with the student's choice of mode. Although equal numbers of students wrote on each version, more Version B students produced alternate modes of writing. Students who wrote on Version A produced 7 alternative mode papers (5 letters and 2 journal entries) while students who wrote on Version B produced 20 alternative mode papers (16 letters, 3 journal entries, and 1 essay). Almost three times as many non-narrative mode papers were produced by students writing on Version B. Most of the alternate mode papers (96.3% or 26 of 27) exhibited personal forms of writing such as letters or journal entries. Most of these more personal forms occurred in response to Version B, the version which more sharply identified an audience for the writer.

There also was a significant relationship between sex and the mode of writing selected by students. ( $\chi^2 = 8.65$ ,  $p < .05$ ). A higher frequency of females (females = 33.3%, males = 10.6%) chose a non-narrative form. In addition, females varied more in their choice of non-narrative modes. Females wrote 15 letters, 5 journal entries, and 1 essay. The five males who chose a non-narrative mode all wrote letters (See Table 7).

However, there was no significant relationship between the mode that was selected and either verbal skill rating ( $X^2 = 8.42$ ,  $p > .05$ ) or attitude rating ( $X^2 = 7.83$ ,  $p > .05$ ). There was also no significant difference in the mean ranking of holistic scores for narrative and non-narrative mode papers ( $X^2_{\text{narrative}} =$

56.40,  $X$  non-narrative = 61.06,  $p > .05$ ). The mean total holistic score for narratives was 6.13, while for non-narratives it was 6.40. Finally, there was no apparent interaction between mode selection (narrative and non-narrative modes) and proficiency when total holistic score was used as the dependent measure. Table 8 indicates that both proficient and less proficient students performed similarly whether they utilized a narrative or a non-narrative mode.

## QUESTIONNAIRE DATA AND COMMENTARY

The following descriptive data are based on the responses obtained on 99 questionnaires. Each data report is followed by a brief commentary.

Memory For The Assigned Prompt. When asked to indicate which version of the prompt they had been assigned for the test, 17 students (14.9%) responded incorrectly; i.e. they thought they had written on a particular version when they actually had not. Most of the students who responded incorrectly (N=13) actually wrote on Version B; eleven who wrote on Version B remembered it as Version A; two who wrote on Version B remembered it as Version C. Only four students who responded incorrectly actually wrote on Version A; all remembered it as Version B.

Commentary: The fact that about 1 in 6 students failed to identify the prompt that they wrote on seems likely to be due to the similar wording in Versions A and B. This interpretation is supported by the observation that most students (N=15) mistakenly identified the prompt they wrote on as either A or B. Only two students mistakenly identified Version C as the topic they wrote on.

The fact that more Version B students thought they wrote on Version A, rather than vice versa, may be due to the order of the versions. Some of Version B students may have selected Version A after reading only the first several lines. The first two sentences of Version A, it should be noted, are identical to the first two sentences of Version B. The final sentence also is identical in both versions.

Task Preference. A majority of the respondents (51.1%) indicated they had no preference in regard to the three versions of the prompt. Of those who had a preference, 39% preferred A, 7% preferred B, and 54% preferred C.

Commentary: One observer suggests that the greater preference for Version C, which provides an elaborate

reality simulation, may simply be a result of the fact that this was the version not written on. This hypothesis could be easily tested by allowing students to choose which of the three versions they wanted to respond to during an actual writing test. It is also possible that students preferred Version C because it sounded least like a writing test prompt. In other words, they thought they would prefer writing for the situation represented by Version C than for the test versions, A and B.

Recollections About The Planning Process.\* A majority of the students (65.7%) indicated they had enough time to complete the task but more than a third of the students (34.3%) reported they did not have enough time to write. Most students (70.7%) indicated they had spent a lot of time (at least 5 or 10 minutes) thinking about what they wanted to write before they started. This is an unusual finding for impromptu timed tests (cf. Pianko, 1979; Emig, 1971) but confirms the observations of the teacher during test administration.\* About a third (34.3%) of the students reported that they felt rushed when they started writing because they spent "so much" time planning. Most students did not re-read their papers. About half (50.5%) indicated they did not have time to re-read, while an additional third 30.3% indicated that although they finished before the time was up, they did not re-read.\*\* Most students (72.7%) indicated they would have written very differently about the prompt if they could have taken the assignment home to complete it, or if they had had several days in class to work on it. A similar number of students (71.1%) reported they could have done a better job on the assignment if they had been given a full hour for writing. Only 20.2% of the students, however, reported that they were anxious while taking the test because it was timed and because they thought it would affect their grade.

\*Responses to the statements in the remainder of the questionnaire are reported with the four response categories collapsed into two: agree and disagree. Percentages of responses in each of the four original categories may be found in Appendix C.

\*The teacher was sufficiently concerned about this long delay that she mentioned it to the investigator before the third testing session, asking, "Shall I tell them to begin writing immediately?"

\*\*This leaves only about 20% of the students who reported both that they finished in time for re-reading and that they actually re-read their papers.

Commentary It appears that this group of writers is less concerned about leaving time for correcting and more concerned about taking time for planning. In using the short time allowance (30-minutes) to best advantage, 70% of the students spent the first 5-10 minutes thinking about what to write. Consequently, some of these planners felt somewhat rushed when they actually started writing. Half of the students did not have the time to re-read, and 30% of the students did have the time but did not re-read.

Responses to the Test Subject. Only 11.1% of the students indicated they wrote better than usual on the timed writing test. A large number (60.6%) reported they had difficulty choosing which one of several possible experiences to write about, indicating the prompt opened access to a wealth of possibilities. However, (34.3%) noted that they had difficulty thinking of any experience to write about, indicating the prompt did not stimulate recall of "first time" memories. A slightly larger proportion (36.7%) indicated they enjoyed the assignment more than they usually enjoyed writing. However, the meaning of this response is ambiguous: It is not clear whether these students liked the prompt better than their usual writing assignments, whether they liked the limited amount of time that they had to spend writing, or whether they responded positively to some other aspect of the test situation. However, 62.2% of the students indicated they would like to write about the prompt when not being tested. This percentage is close to the number of students who felt they would have treated the prompt differently if they could have done the writing at home.

Most of the students (78.8%) indicated that they had never written about a first time experience before. More than half (54.5%), however, indicated that they had talked with others about the particular experience they had chosen to write about.

Commentary. This latter finding supports others described elsewhere. It suggests that this invitation to recall a first time experience encourages students to draw on anecdotes that are readily available in memory and may already have been rehearsed orally. It would be useful to investigate how the properties of different prompts either function to constrain or to enable different assessment populations in selecting possible experiences to write about.

Writers' Self-Profiles: Other Writing Experiences. A majority of the students considered themselves good writers (65.7%). Only 16.2% indicated that

they preferred writing academic essays or non-fiction reports rather than short stories or personal essays about their own experiences. While most of the students (68.7%) did not keep a diary or journal of experiences or thoughts, nearly half (48.5%) indicated that they sometimes wrote stories, poems, or other pieces outside of school when the writing was not assigned. About a third of the students (28.6%) indicated they did not have much prior experience with timed writing tests. A majority of students (67.7%) did not think the test was a good measure of their abilities as writers. Only 39.4% thought the test differed from other timed writing tests they had taken. In addition, 68.6% of the students thought the test required the kind of writing they had to do for other high school English assignments.

Commentary: These observations suggest two possible interpretations of the way students perceived the task in relation to their school writing assignments. First, the students may have felt that both this timed writing test and their typical school assignments are poor measures for assessing their writing ability. Specifically, students may have thought that this test prompt did not provide a good measure of their writing ability and, because the test prompt was similar to other assignments they receive in high school, school writing assignments are also poor measures of their writing ability. Alternatively, the students may have been responding to the timed aspect of the test situation when they reported that it was a poor measure of their writing ability. And they may have been responding to the subject when they reported that the test required the kind of writing required for other high school English assignments. If they were responding to the timed aspect of the test, the students did not mean to give a pejorative evaluation of their high school assignments when they reported that this test was not a good measure of ability.

The latter interpretation seems more acceptable given the fact that a majority of the students (72.7%) believed they would have written very differently, and would have done a better job, if they had been allowed more time to write. The fact that a majority of the students (62.2%) indicated they would like to write about the prompt when not being tested lends additional support to the latter interpretation. Finally, the teacher reported that the prompt for this timed writing test, calling for expressive writing from personal experience, was similar to several other classroom assignments given to this particular population of students. The available data suggests, then, that the students were responding to the nature of the writing

occasion (a timed test) rather than to the "kind" of writing required by the test when they called the test a poor measure of their ability.

251

## RELATIONSHIPS BETWEEN SELECTED VARIABLES AND SPECIFIC QUESTIONNAIRE ITEMS

When responses to specific questionnaire items were examined in relation to variables such as verbal skill, attitude, sex, proficiency, and holistic score, additional interesting information was found.

### Verbal Skill Rating and Questionnaire Items

With one exception, no significant relationships were found between students' verbal skill rating and individual questionnaire items. The only significant relationship found was between verbal skill rating and Item 18 ( $Tau_c = .23, p < .05$ ). As the teacher's rating of the student's verbal skill increased, students tended to agree with the statement, "I consider myself a good writer." Thus, student self-perception tended to agree with teacher perception.

### Attitude Rating and Questionnaire Items

Significant positive relationships were found between the teacher's rating of student attitude toward academic achievement and several questionnaire items. As attitude rating increased there was greater agreement with the following statements:

3. I did not have enough time to write.  
( $Tau_c = .26, p < .05$ )
8. I did not have time to re-read or proofread my paper.  
( $Tau_c = .20, p < .05$ )
10. I believe I would have written very differently about this topic if I could have taken the assignment home to complete, or if I had several days to work on it in class.  
( $Tau_c = .17, p < .05$ )
11. I believe I could have done a better job with this assignment if I had been given an hour to write.  
( $Tau_c = .27, p < .05$ )

Significant negative relationships also were found between attitude and several questionnaire items. As attitude rating decreased, there was greater agreement with the following items:

9. Although I finished before the time was up, I did not re-read my paper.  
( $Tau_c = -.18, p < .05$ )
24. I think the timed writing test we took is a good measure of my abilities as a writer.  
( $Tau_c = -.23, p < .05$ )

In general, then, those students who are perceived by the teacher as being conscientious about their academic achievement and positive about their work are the ones who felt the 30-minute test did not allow them enough time to demonstrate their usual or best level of performance. These are students who would re-read, given time, and who are likely to have set more complex goals for themselves had they been working at home with unlimited time. Students with poorer attitudes toward achieving in school did less re-reading, even if there was time, and were generally satisfied that the timed test adequately measured their abilities.

#### Sex and Questionnaire Items\*

Males and females appeared to differ in response to the time demands of the task. Although most students indicated they had enough time to finish the task, females tended to indicate the opposite more frequently than males; 43.9% of the females reported that they did not have enough time to write, compared with only 21.4% of the males. This association between sex and response to item 3 was significant ( $\chi^2_{sup 2} = 4.45, p < .05$ ).

Significant associations were also found between sex and responses to questionnaire item 20 ("I keep a diary or journal of my experiences or thoughts.") and item 21 ("I sometimes write stories or poems or other pieces outside of school, even though the writing was not assigned."). A journal or a diary was kept by 43.9% of females, but only 14.3% of males reported that they kept a diary or journal ( $\chi^2 = 8.51, p < .05$ ). Similarly, 57.9% of females but only 35.7% of males reported that they wrote unassigned pieces outside of school ( $\chi^2 = 3.92, p < .05$ ).

\*This analysis was conducted after collapsing the four categories of responses on the questionnaire into two categories (disagree and agree).

## Proficiency and Questionnaire Items\*

When the questionnaire data were examined in terms of the proficiency level of the students, several very striking and consistent patterns appeared. First, there were significant associations between proficiency level and responses to questionnaire items 3, 4, 6, 7, 8, and 9. These associations seem to suggest that proficient writers 1) used a substantial amount of planning time, 2) felt rushed when they actually started writing, 3) felt they didn't have enough time to write, and 4) did not have time to reread their papers.

There was a significant relationship ( $\chi^2 = 4.51$ ,  $p < .05$ ) between proficiency and response to the statement: "I spent a lot of time (at least five or ten minutes) thinking about what I wanted to write before I started." A majority of the proficient writers (77.0%) agreed with this statement compared to 52.0% of the less proficient writers. At least part of the additional planning time that proficient writers seem to spend was taken up in deciding which of several possible experiences they should write about. For example, 67.6% of the proficient writers, compared to 40.0% of the less proficient writers, agreed with the statement: "I had a hard time choosing which one of several possible experiences I should write about." ( $\chi^2 = 4.85$ ,  $p < .05$ ). Writing proficiency seems to be associated with taking 5-10 minutes of planning time to consider a range of alternatives accessible in memory in selecting a specific experience to write about.

Because proficient writers take so much time to plan their composition, they often feel rushed when they finally commence writing. Proficiency was significantly associated with agreement with item 7, "Because I spent so much time planning, I felt rushed when I actually started writing." ( $\chi^2 = 8.79$ ,  $p < .05$ ). This association is suggested by the fact that 43.2% of proficient writers agreed with this statement compared to only 8% of the less proficient writers.

The extended planning time typical among these proficient writers may also contribute to their feeling about the amount of time necessary to adequately complete the assignment. Proficiency was significantly associated with agreement with item 3: "I did not have enough time to write." ( $\chi^2 = 8.79$ ,  $p < .05$ ).

---

\*This analysis was conducted after collapsing the four categories of responses on the questionnaire into two categories (disagree and agree).

Proficient writers more frequently agreed with item 3 compared to less proficient writers (Proficient = 43.2%, Less Proficient = 8.0%).

Finally, the extended planning time typical of proficient writers may have prevented many of them from re-reading or proof-reading their papers. There was a significant association between proficiency and agreement with item 8: "I did not have time to re-read or proof-read my paper." ( $\chi^2 = 8.03, p < .05$ ). Some 59.5% of proficient writers agreed with this statement compared to 24.0% of less proficient writers.

There is also some indication that less proficient writers may be less likely to re-read their papers. For example, 52.0% of the less proficient writers agreed with item 9, "Although I finished before the time was up, I did not re-read my paper.", compared to only 23.0% of the proficient writers ( $\chi^2 = 6.14, p < .05$ ). This association may, however, be due to the greater number of less proficient writers who had sufficient time to re-read their papers (less proficient = 76.0%, proficient = 40.5%).

In addition to the associations that exist between proficiency, planning time, and the consequences of taking 5-10 minutes to plan one's composition during a 30-minute test period, there was an association between proficiency and student's attitudes about how additional time to write would have influenced their writing performance. Proficiency was positively associated ( $\chi^2 = 5.91, p < .05$ ) with the belief that the final product would have been very different if students had several days to complete the task. Of the proficient writers 79.7% agreed with this statement (item 10) compared to only 18.1% of the less proficient writers. Proficiency was similarly associated ( $\chi^2 = 18.75, p < .05$ ) with the belief that the final product would have been better if students had had an hour to write. Of the proficient writers 83.8% agreed with this statement (Item 11) compared to 36.0% of the less proficient writers. Proficient writers believed that additional time would have changed and benefited their writing performance.

Commentary: The pattern of results from investigating the relationship between proficiency and questionnaire items suggests that a 30-minute timed writing test may not adequately measure the ability of proficient writers. Placing a very short time limit on the writing task may create a ceiling effect on students' performance. Even though their writing was generally superior, proficient writers report that they could have done even better if they had had more time. The fact

competent writers. Further empirical support for this first finding in additional studies would suggest that short timed writing tests are likely to severely truncate the range of performance elicited during the assessment, although good writers will still receive top scores on a short test. The second finding raises some question about whether the students in this study received high holistic scores because they were generally better writers or because they had previously written about the topic. It may, at a later date, be profitable to explore the effects of prior experience with a topic on writing performance. The third finding indicates simply that good writers are aware of their abilities. It does, however, raise some question about the effects of self-evaluation on writing ability, i.e. does a positive concept of oneself as a writer facilitate writing performance?

#### Prompt Version and Mode and Planning Time Questionnaire Items

Table 9 summarizes the results of the questionnaire data by prompt version. Many items were not associated with differences between versions. Differences did appear, however, in several response items, and this suggests an interesting effect of prompts on attitudes toward the test experience as well as an interesting effect of prompts on the writing process itself.

For example, more students who wrote on Prompt Version B seemed to think the test was a good measure of their abilities as a writer (Item 24). Of the students who wrote on Prompt Version B, 40.8% agreed with this statement compared to only 24.5% of students who wrote on Prompt Version A.

All of the other important differences between versions were related to the way students described their writing process (Items 4, 6, and 7). A higher frequency of students who wrote on Version A indicated that they spent a significant amount of time planning their response. More students who wrote on Version A, for example, reported that they spent at least 5-10 minutes thinking about what they wanted to write (Version A = 79.6%, Version B = 62.0%). More students who wrote on Version A also reported that they had difficulty choosing which of several possible experiences to write about (Version A = 75.5%, Version B = 46.0%). The apparent discrepancy in amount of planning time between prompt versions is also supported by students' response to Item 7. More students who wrote on Version A felt rushed when they actually started writing (Version A = 40.8%, Version B = 28.0%)

In general, students who wrote for a less specified audience (Version A) more frequently spent a substantial amount of time planning what they were going to write. Students who wrote for a more specified audience (Version B) less frequently spent a substantial amount of time planning. This difference is illustrated in Figure 1.

It is important to note, however, that despite these process differences there was no significant difference between versions in the final quality of the product as measured by total holistic score (Mann-Whitney  $U = 1480.5$ ,  $z = -.85$ ,  $p > .05$ ). Of the 99 students that completed both the writing test and the questionnaire, the mean holistic score for Version A students was 6.09 ( $N=49$ ); for Version B students the mean holistic score was 6.28 ( $N=50$ ). On the surface, at least, these versions appear to be similar with respect to the holistic scores that they produce.

In addition to an analysis of the main effects, an analysis of the interaction between prompt version and planning time on writing performance was conducted. Table 10 indicates that students who wrote on Version A and students who wrote on Version B also performed somewhat similarly.

In addition to this two-way analysis between prompt version and planning time, a three-way analysis was conducted between prompt version, planning time, and re-reading. Table 11 shows the 8 possible combinations of these three variables, the means and variances of the holistic scores for each combination, and the number of students that fell into each group.

In the writing performance of Version A students, re-reading seemed to interact with planning time. Version A students in Group 4 who re-read their papers and spent 5-10 minutes planning had a higher mean holistic score ( $X = 6.42$ ) than students in Group 2 who re-read their papers but did not plan ( $X = 5.67$ ). Apparently, re-reading only had an effect for Version A students if, in addition, they spent 5-10 minutes planning.

Not taking time for re-reading, among Version A students, also interacted with planning but in the opposite direction. That is, Version A students in Group 1 who did not re-read their papers and who did not spend 5-10 minutes planning had higher mean holistic scores ( $X = 6.43$ ) than students in Group 3 who did not re-read and did plan ( $X = 5.93$ ). Not re-reading, for Version A students, was associated with higher holistic scores only for those students who also did not plan.

Among Version B students, re-reading also interacted with planning time. Students in Group 6 who re-read and did not plan had higher mean holistic scores ( $X = 6.29$ ) than students in Group 8 who both re-read and planned ( $X = 6.00$ ). Apparently, re-reading only had a positive effect for Version B students if, in addition, they did not plan.

Not re-reading, among Version B students, also interacted with planning time. Version B students in Group 7 who planned but did not re-read had higher mean holistic scores ( $X = 6.55$ ) than students in Group 5 who neither planned nor re-read ( $X = 6.25$ ).

Additional findings of interest between prompt version and questionnaire items involve the relationships that were found with the mode of writing that students selected. More students who wrote narratives (73.7%) and journals (75.0%) indicated that they spent 5-10 minutes planning than students who wrote letters (55.6%). More students writing in non-narrative forms finished and re-read their papers (letter = 83.3%, journal and essay = 100%) than did those students writing a narrative (64.5%). Overall, those students who wrote letters generally felt they could have done a better job with more time (88.9%) compared to those who wrote narratives (68.4%) or journal entries (50.0%).

Regarding opinions about the writing task, all of the students choosing non-narrative forms ( $N=23$ ) did not think they wrote better on this assignment than they usually do. This compares with only 14.4% of the students who chose a narrative mode.

These response frequencies suggest that selecting a non-narrative mode, particularly the letter form, had the effect of making these students wonder whether the task enabled them to show their true ability.

As reported elsewhere, although students generally thought that the task was not a good measure of their writing ability, letter writers reported this more often than did students who wrote narratives or journals (letter = 77.8%, narrative = 67.1%, journal = 50.0%).

## DISCUSSION

One of the more interesting results from this study is the observation that differences in prompt version did not produce a difference in performance as measured by total holistic score means. The mean performance of students who wrote on Version A was nearly identical to the mean performance of students who wrote on Version B. This would seem to suggest that increasing the degree to which the writer's audience and rhetorical purpose are specified has no effect on the quality of a timed writing performance. There are several possible explanations for this result, however, that must be considered. Each raises an additional research issue that should be explored before concluding that audience specification has no effect on the quality of a student's performance.

It should be noted, first, that the two versions are not extreme examples; that is, they are not at opposite ends of an audience specification continuum. In Version A, the audience, though general, is nevertheless mentioned (i.e., "the reader"). Moreover, including a clearly specified communicative purpose (i.e., "Write about this 'first time' so that your reader understands how you felt and why this experience stayed in your memory") is likely, by itself, to constrain the set of possible audiences and the sense of rhetorical purpose. Implicit in this task is the idea that one should write to an audience that can understand and appreciate a writer's particular "first time" experience. In Version B, the audience is only vaguely specified (i.e., "Imagine that you are writing to someone who has just had a similar experience, or to someone who is about to have such an experience") The writer does know something about the reader; he or she has or will have a similar experience, but no more than that. A much more detailed specification of the audience is clearly possible. The writer could have been asked, for example, to explain the experience to a retired man, in his sixties, who was recently widowed and who will shortly share a similar experience. Clearly what is needed is additional research on pairs of prompts in which the difference in audience specification is made more evident. The failure to find differences in the quality of writing between these two versions may have been due to the fact that they both represent middle range examples on the continuum of audience specification.

Though there was no association evident between the assigned prompt and holistic score, the effects of the prompts are clearly associated with the students' choice of mode. Equal numbers of students wrote on

each version, but Version B students produced more alternate modes of writing. These students produced 20 alternate mode papers (16 letters, 3 journal entries, 1 essay), while students who wrote on Version A produced only 7 alternate mode papers (5 letters, 2 journal entries). Most of the more personal forms occurred in response to Version B.

Although mean holistic scores did not significantly differ, there were clear differences in the planning process. Decreasing the audience specification seemed to produce an increase in planning time. It may be that a topic lacking both implicit and explicit audience specification may produce even greater differences in planning time, at least among writers who are willing and able to use time to plan during a test requiring impromptu writing.

There is a second possible explanation for the failure to find differences in the quality of the writing that students produced in response to the two prompt versions. As mentioned earlier, the sample population differed in certain characteristics from what would be expected in a fully representative sample of high school students. The distribution of scores for verbal skill, attitude, and holistic score are all skewed towards the high end of the scale. For example, 31.5% of the papers, received one of the two highest holistic scores (7-8). Only 5.4% of the papers received one of the two lowest scores (2-3). Because scores on the performance measure were skewed high, a topping out effect may have minimized the range of differences and therefore limited the likelihood of finding a significant difference between the mean holistic scores for the two versions. A student population in which holistic scores are distributed more normally or a reading using Diedrich's (1974) method of forced quartile distribution may be more likely to produce significant differences in prompt effect.

A third explanation for equivalent mean scores for the two prompt versions is compatible with the findings of complex interactions between various composing strategies, including planning and re-reading. The difference in number as well as in the product performance of planners and non-planners in the two version groups may have generated similar mean scores for the two versions, although in fact, the prompt versions differentially affected the product quality of individuals. With further research it should be possible to determine whether the processes chosen were natural to the students who chose them, or whether they were influenced by the prompt version, and whether these were optimal process decisions for successful performance.

Observing the same students over several writing tasks would reveal whether some kinds of tasks consistently elicit or require more planning time than others, at least among students who plan, and whether planning consistently produced better products.

We may finally speculate on the effect of the observation that Version A, because of fewer constraints, seems less likely to quickly provide students with a good subject to write about. Students who are serious about their writing and who therefore seek subjects they care about, may be required to spend more time searching when faced with Version A. (Less serious, more perfunctory writers faced with Version A may begin writing quickly, having fewer resources for scanning and selecting appropriate subjects -- hence the association of lower scores with Version A non-planners.) Version B, on the other hand, while allowing such a search, seemed to trigger more instant responses in a wider range of students. Students appeared to quickly choose a subject which usually engendered successful writing within the given time limit. The effect of this difference may be paradoxical: the extra search for meaningful, appropriate subject matter engaged in by Version A students may have contributed something to the composing process that took these papers to the highest level of quality -- a level less available to Version B students whose subject came more quickly to mind.

The implications of such a finding, if it were to be confirmed by more precise observations of the composing process and analysis of the written products, would be minimal for general psychometric purposes, but would be important for the evaluation of instructional programs and for research measures which seek to assess differences in the performance of able as well as less able student writers. The large difference in the frequency with which different rhetorical modes were produced by the two prompt versions is also of interest. The version with the greatest audience constraint (B) produced far more papers of a personalized type (journal entry or letter). It would be useful to know if this is due solely to increased audience specification or if it is also a function of the particular communicative purpose implied in Version B. Explaining a first time experience to someone who is about to or has just had a similar experience the writer shares an important adventure with his or her reader. It may, thus, call for a shift in the register or the form selected by the writer. It would be interesting to know if more personalized forms are also produced when the audience specified excludes peers: "Explain this first-time experience to a prospective employee who has

just had or is about to have a similar experience."

Another finding of this study seems also to require additional investigation. The fact that a substantial proportion of these students (72.7%) reported that they would have written very differently about the prompt if they had had more time to work on it raises questions about the ecological validity of holistic scores on a single sample as an assessment measure. It suggests that students' performance on a short, timed writing test may differ substantially from the conditions more typically associated with regular classroom writing assignments, at least in this teacher's class. It would be useful to understand how performance differs as the amount of available writing time increases: are different forms of writing produced when students have more time? Does the quality of writing differ substantially? What type of student seems to revise the most when greater time is available? How is the writing process different when greater time is available? Which individual differences (grade, sex, verbal ability, attitude, etc.) seem to interact with differences in the amount of available writing time and seem to produce changes in the quality of writing?

In general, students who are likely to receive top scores on longer tests also receive top scores on short tests. Short, timed tests, then, may be adequate in a gross way to identify proficient students or predict success at longer tasks. However, they may be inadequate to distinguish among various levels of proficiency and hence may be poor measures for evaluating writing programs which aim to move writers along a scale toward greater and greater proficiency.

If the purpose of an assessment is merely to measure minimum proficiency, short tests may provide a sufficient range of performance among most populations. But since short tests may reveal only the grossest differences in ability among homogeneous populations, improvements in writing ability over a semester or year course as a result of certain instructional practices may be totally obscured, with students appearing to make little or no improvement. Hence, short tests may be poor research or evaluation measures for improving instruction.

It may be argued that writing programs are not making important enough differences unless the differences can be recorded by fairly rough measures which distinguish between proficient and non-proficient writing -- and hence, that short timed tests are sufficient measures of program effect. But this is tantamount to saying that the only growth in writing that matters is

the movement from an inability to the ability to produce minimally competent writing during a 30-minute writing task. This puts a ceiling not only on our measure but on our goals for writing instruction.

We must keep in mind, of course, that all of the above observations are based on reports by students of their own writing processes. Although the teacher observation that the class was taking longer than usual supports the self-reports, the above results should be viewed only as suggestive of questions which might be best answered not by questionnaire, but by direct observation of individual students during the act of composing. Many such observations will be needed to determine whether the kind of trend noted here is a significant one, and what its implications are for test design and for our understanding of the composing processes of writers on impromptu timed writing tests.

Because statistical tests of significance were employed in analyzing the data collected for this study, it should not be concluded that the researchers are committed only to hypothesis testing. Rather, we are convinced that the current state of understanding about the performance of students on timed writing tests is sufficiently primitive that a wide range of general observations needs to be made in the hope of generating testable hypothesis, which in turn will lead to a refinement of our measures. Many conclusions about student writing ability and the effect of different instructional programs are currently based on measures whose accuracy and power are subject to serious question. We believe the data of this preliminary study, while suggesting few clear answers to our initial questions, strongly confirm that those questions are worth asking and are potentially answerable.

Table 1

The Student Population Used In This Study  
by Grade Level, Sex, Subject and Section

	Advanced Comp. Section 1 Jrs. & Srs.	English Lit. Sect. 1    Sect. 2 Jrs.        Jrs.		General English Sect. 1 Soph.	Total
<b>Males</b>					
N	12	12	13	11	48
%	(10.7)	(10.5)	(11.4)	(9.6)	(42.1)
<b>Females</b>					
N	10	21	19	16	65
%	(3.8)	(18.4)	(16.7)	(14.0)	(57.9)
<b>Column Total</b>					
N	22	33	32	27	114
%	(19.3)	(28.9)	(28.1)	(23.7)	(100.0)

Table 2

Distribution of Verbal Skill Ratings

	<u>Well Below Grade Level</u>	<u>Grade Level or Below</u>	<u>Grade Level or Above</u>	<u>Well Above Grade Level</u>	<u>Missing Cases</u>	<u>Total</u>
N	3	39	53	17	2	114
%	(2.6)	(34.2)	(46.5)	(14.9)	(1.8)	(100.0)

Table 3

Distribution of Attitude Ratings

	<u>Extremely Negative</u>	<u>Mildly Negative</u>	<u>Mildly Positive</u>	<u>Extremely Positive</u>	<u>Missing Cases</u>	<u>Total</u>
N	9	25	44	34	2	114
%	(7.9)	(21.9)	(38.6)	(29.8)	(1.8)	(100.0)

Table 4

Distribution of Total Holistic Scores

	2	3	4	5	6	7	8	Total
N	0	2	13	10	47	21	21	114
%	(0.0)	(1.8)	(11.4)	(8.8)	(41.2)	(18.4)	(18.4)	(100.0)

Table 5

Means and Standard Deviations of Proficient Students' and Less Proficient Students' Total Holistic Scores, by Class, Section and Grade Level

	PROFICIENT STUDENTS				LESS
	Advanced Comp. Section 1 <u>Jrs. &amp; Srs.</u>	English Lit. Section 1 <u>Jrs.</u>	Section 2 <u>Jrs.</u>	<u>Total</u>	PROFICIENT STUDENTS General English Section 1 <u>Soph.</u>
$\bar{X}$	6.64	6.49	6.41	6.50	5.19
S.D.	1.049	1.25	1.13	1.21	1.11
N	22	33	32	87	27

Table 6

Frequencies for the Different Modes of Writing Produced by Prompt A and Prompt B (Given in Percentages)

	Narrative	Letter	Journal	Essay	Raw Total
Prompt A					
%	87.7	8.8	3.5	0.0	50.0
N	(50)	(5)	(2)	(0)	(57)
Prompt B					
%	54.9	28.1	5.3	1.8	50.0
N	(37)	(16)	(3)	(1)	(57)
Total					
%	76.3	18.4	4.4	0.9	100.0
N	(87)	(21)	(5)	(1)	(114)

Table 7

Sex by Mode Frequencies

	Narrative Mode	Non-Narrative Mode			Total
		Letter	Journal	Essay	
Females					
$\bar{X}$	66.7	24.2	7.6	1.5	57.9
N	(44)	(16)	(5)	(1)	(66)
Males					
$\bar{X}$	89.5	10.6	0.0	0.01	42.1
N	(43)	(5)	(0)	(0)	(48)

Table 8

Mean Total Holistic Scores by Proficiency and Mode Selection

	<u>Proficient Ss</u>	<u>Less Proficient Ss</u>
<u>Narrative Mode</u>		
$\bar{X}$	6.48	5.09
S.D.	1.11	1.19
N	65	22
<u>Non-Narrative Mode</u>		
$\bar{X}$	6.55	5.60
S.D.	1.30	0.55
N	22	5

Table 9

Frequency of Student Responses to Questionnaire Items  
3-25 by Prompt Version

ITEM No.	VERSION A (N=49)		VERSION B (N=50)	
	Disagree	Agree	Disagree	Agree
3.	65.3	34.7	66.0	34.0
4.	24.5	75.5	54.0	46.0
5.	63.3	36.7	68.0	32.0
6.	20.4	79.6	33.0	67.0
7.	59.2	40.8	72.0	28.0
8.	53.1	46.9	46.0	54.0
9.	63.3	36.7	76.0	24.0
10.	28.6	71.4	26.0	74.0
11.	32.7	67.3	24.0	76.0
12.	77.6	22.4	82.0	18.0
13.	89.8	10.2	89.0	12.0
14.	64.6	35.4	62.0	38.0
15.	39.6	60.4	36.0	64.0
16.	75.5	24.5	92.0	8.0
17.	42.9	57.1	48.0	52.0
18.	36.7	63.3	32.0	68.0
19.	81.6	18.4	96.0	4.0
20.	65.3	34.7	72.0	28.0
21.	49.0	51.0	54.0	46.0
22.	68.8	31.3	74.0	26.0
23.	57.1	42.9	64.0	36.0
24.	75.5	24.5	60.0	40.0
25.	26.5	73.5	34.0	66.0

Table 10

Total Holistic Score Means, Standard Deviations  
and Numbers of Students by Prompt and Planning Time

	Prompt A	Prompt B
Spent 5-10 Minutes Planning		
$\bar{X}$	6.03	6.39
S.D.	1.31	1.28
N	39	31
Did Not Spend 5-10 Minutes Planning		
$\bar{X}$	6.20	6.26
S.D.	1.23	1.20
N	10	19

267

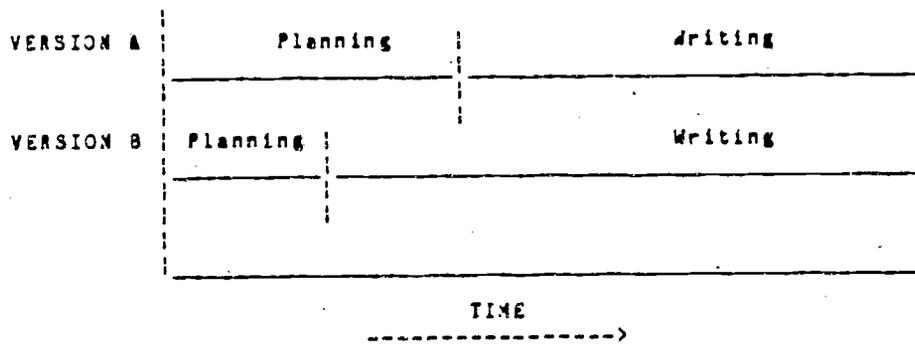
Table 11

Total Holistic Score Means and Standard Deviation  
by Prompt, Planning Time and Re-Reading

<u>VERSION A</u>	<u>Group 1</u>	<u>Group 2</u>	<u>Group 3</u>	<u>Group 4</u>	
Greater Audience Constraint	No	No	No	No	
Spent 5-10 Minutes Planning	No	No	Yes	Yes	
Finished and Re-read	No	Yes	No	Yes	
	$\bar{X}$	6.43	5.67	5.93	6.42
	S.D.	1.13	1.53	1.24	1.44
	N	7	3	27	12
<u>VERSION B</u>	<u>Group 5</u>	<u>Group 6</u>	<u>Group 7</u>	<u>Group 8</u>	
Greater Audience Constraint	Yes	Yes	Yes	Yes	
Spent 5-10 Minutes Planning	No	No	Yes	Yes	
Finished and Re-Read	No	Yes	No	Yes	
	$\bar{X}$	6.25	6.29	6.55	6.00
	S.D.	1.10	1.50	1.18	1.50
	N	12	7	22	9

Figure 1

A Schematic Diagram of the Differences in Planning and Writing Time for Versions A and B



269

## REFERENCES

- Cooper, C.R. & Odell, L. Evaluating Writing: Describing, Measuring, Judging. Urbana, Ill.: NCTE, 1977.
- Diederich, P.B. Measuring Growth in English. Urbana, Ill.: NCTE, 1974.
- Emig, J. The Composing Processes of Twelfth Graders. Urbana, Ill.: NCTE Research Report No. 13, 1971.
- Pianko, S. A description of the composing process of college freshmen. Research in the Teaching of English, 1979, 13(1), 5-22.
- Woodworth, Patrick and Keech, Catharine. The Write Occasion. Bay Area Writing Project, University of California, Berkeley, 1980.

APPENDIX A: WRITING TEST

STUDENT NAME \_\_\_\_\_ TOPIC \_\_\_\_\_  
(as you normally write it on English papers)

Student Code:                        

Instructions to Students:

1. Please write your name in the appropriate blank.
2. Look briefly at the test sheet attached. You have either Topic 1 or Topic 2. Write the number of your topic in the space above beside the word "Topic."
3. To fill in your student code, please put the following information in the appropriate blanks above:
  - blank 1: The period you have English with Ms. McKelly (3,4,5, or 6).
  - blanks 2 - 5: Please use the last four digits of your phone number, of four digits of your house number, whichever you prefer, to be your own individual code number.
  - blank 6: Please write the number 1 if you are female; please write the number 2 if you are male.
  - blank 7 and 8: Please leave these spaces empty. If you copy your code onto additional sheets of paper, please leave room for these two digits to be added by your teacher.
4. Once you have created your code, minus the last two digits, please copy it into the appropriate blanks on the attached test sheet. DO NOT WRITE YOUR NAME on the test sheet. Your code will assure that your paper is returned to you, but your reader will not know who you are while he or she is scoring your paper.
5. You may write as much as you wish in the time available. (30 minutes.) If you need more paper, please be sure to write your code number in the top righthand corner of each page you write on.
6. Please use pen or dark pencil so your paper can be duplicated if necessary.
7. If you finish writing before the time is up, please read quietly or look out the window or stare into space, so that others who are not done can continue writing without distraction.
8. THANK YOU FOR YOUR PARTICIPATION IN OUR WRITING SAMPLE! To do best in this writing experience! You will do best if you concentrate on what you have to say rather than on trying not to make errors in your writing. Your readers will know that you had only 30 minutes to write and little time for editing.

## APPENDIX A: WRITING PROMPTS

Students were given one of the following two versions of the prompt.

### VERSION A

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Write about this "first time" experience so that your reader can understand your feelings and why this memory has stayed with you.

You may write an essay, a short story, a letter, or a journal entry as a way of re-telling your experience.

### VERSION B

Think of an occasion when you experienced something for the FIRST TIME. It may have been something you later came to do more easily or something you now take for granted; but, for some reason, your first time was memorable. Imagine that you are writing to someone who has just had a similar experience, OR to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or it might help your reader understand that other people have gone through the same kind of thing.

You may write an essay, a short story, a letter, or a journal entry as a way of re-telling your experience.

## APPENDIX B: QUESTIONNAIRE

### Writing Test Questionnaire.

Please read and compare the following three versions of the writing idea you used recently for a writing test. Then answer the questions which follow.

---

A. Think of an occasion when you experienced something for the FIRST TIME. It may be something you later came to do easily or something you now take for granted; but, for some reason the first time was memorable. Write about this "first time" so that your reader understands how you felt and why this experience stayed in your memory.

You may write an essay, a short story, a letter, or a journal as a way of re-telling your experience.

B. Think of an occasion when you experienced something for the FIRST TIME. It may be something you later came to do easily or something you now take for granted; but, for some reason, the first time was memorable. Imagine you are writing to someone who has just had a similar experience, or to someone who is about to have such an experience. Your writing might help prepare your reader for the experience, or help your reader understand that other people have gone through the same kind of thing.

You may write an essay, a short story, a letter, or a journal as a way of re-telling your experience.

C. There is a popular book called MY FIRST TIME, edited by Bennett Cerf, which is a collection of pieces by famous people who are describing experiences they remember when they were doing something for the first time. They tell about first kisses, first time on stage, first crime, first time they saw someone famous they would later work with, first political campaign, first time they were afraid, etc. Some of their experiences are unusual, some are common but are well-told.

I would like to collect a book of FIRST TIME experiences of high school students. I have asked a number of teachers to allow their students to write about their FIRST TIMES during a 30 minute writing test, from which I will select promising pieces. Later I will contact the authors of the selected pieces and give them an opportunity to revise and polish before publication. Chosen pieces will be copyrighted and the authors will receive \$50 for their contribution. No piece will be used without the author's permission.

You may write an essay, a short story, a letter, or a journal entry as a way of re-telling your experience.

---

I. (cont.) Please answer the following questions about the three versions of your writing topic given above:

- |                                |   |   |   |                |
|--------------------------------|---|---|---|----------------|
| 1. I wrote on topic.....       | A | B | C | don't remember |
| 2. I would prefer to write on: | A | B | C | n. preference  |

---

Please explain or comment if you have a preference:

(continue on the reverse side)

II. Circle appropriate number in the following responses:)

1 = Strongly disagree    2 = Disagree    3 = Agree    4 = Strongly agree

- |   |   |   |   |   |
|---|---|---|---|---|
| 3. I did not have enough time to write  | 1 | 2 | 3 | 4 |
| 4. I had a hard time choosing <u>which</u><br><u>one</u> of several possible experiences<br>I should write about.   | 1 | 2 | 3 | 4 |
| 5. I had a hard time thinking of ANY<br>experience worth writing about.   | 1 | 2 | 3 | 4 |
| 6. I spent a lot of time (at least<br>five or ten minutes) thinking<br>about what I wanted to write<br>before I started.  | 1 | 2 | 3 | 4 |
| 7. Because I spent so much time<br>planning, I felt rushed when I<br>actually started writing.  | 1 | 2 | 3 | 4 |
| 8. I did not have time to re-read<br>or proof-read my paper.  | 1 | 2 | 3 | 4 |
| 9. Although I finished before the<br>time was up, I did not re-read my<br>paper.  | 1 | 2 | 3 | 4 |
| 10. I believe I would have written<br>very differently about this topic<br>if I could have taken the assign-<br>ment home to complete, or if I<br>had several days to work on it in<br>class. | 1 | 2 | 3 | 4 |
| 11. I believe I could have done a<br>better job with this assignment<br>if I had been given an hour to write.   | 1 | 2 | 3 | 4 |

Page 3  
Writing Questionnaire

- |   |   |   |   |   |
|---|---|---|---|---|
| 12. I was anxious while I was writing because this was a timed writing test which would affect my grade.      | 1 | 2 | 3 | 4 |
| 13. I think I wrote better for this assignment than I usually write.  | 1 | 2 | 3 | 4 |
| 14. I enjoyed this assignment more than I usually enjoy writing.  | 1 | 2 | 3 | 4 |
| 15. I would like to write about this topic when I am not being tested.  | 1 | 2 | 3 | 4 |
| 16. I have written about this topic (a first time experience) before, as a school assignment.                 | 1 | 2 | 3 | 4 |
| 17. The particular experience I chose to write about is one I have talked about or told someone about before. | 1 | 2 | 3 | 4 |

---

III. Please continue circling the response that comes closest to describing your feelings about yourself:

- |  |   |   |   |   |
|--|---|---|---|---|
| 18. I consider myself a good writer.   | 1 | 2 | 3 | 4 |
| 19. I prefer writing academic essays or non-fiction reports rather than short stories or personal essays about my own experiences. | 1 | 2 | 3 | 4 |
| 20. I keep a diary or journal of my experiences or thoughts.   | 1 | 2 | 3 | 4 |
| 21. I sometimes write stories or poems or other pieces outside of school, even though the writing is not assigned.                 | 1 | 2 | 3 | 4 |
| 22. I have not had much experience with timed writing tests.   | 1 | 2 | 3 | 4 |
| 23. The timed writing test we took for this experiment was different from other timed writing tests I have had.                    | 1 | 2 | 3 | 4 |

---

(Please explain your answer to 23 if you agreed or strongly agreed:  
Continue on the reverse side of this paper.)

Page 4  
Writing Questionnaire

24. I think the timed writing test we took is a good measure of my abilities as a writer.

1 2 3 4

25. I think the test we took requires the kind of writing I must be able to do for other high school English assignments.

1 2 3 4

2/11

APPENDIX C

FREQUENCY OF RESPONSES FOR QUESTIONNAIRE ITEMS 3 - 25

Item No.	Strongly Disagree	Disagree	Agree	Strongly Agree
3.	0.1	45.6	25.7	28.6
4.	11.1	29.3	49.5	11.1
5.	21.2	41.1	31.2	17.4
6.	15.4	24.2	51.8	18.6
7.	11.1	31.5	31.2	26.2
8.	3.4	42.1	39.1	15.4
9.	27.2	45.5	25.7	11.6
10.	1.1	21.2	31.7	46.1
11.	15.4	27.2	42.1	15.3
12.	11.1	31.2	31.1	26.6
13.	15.4	31.2	31.1	22.3
14.	11.1	31.2	31.1	26.6
15.	11.1	31.2	31.1	26.6
16.	11.1	31.2	31.1	26.6
17.	11.1	31.2	31.1	26.6
18.	11.1	31.2	31.1	26.6
19.	11.1	31.2	31.1	26.6
20.	11.1	31.2	31.1	26.6
21.	11.1	31.2	31.1	26.6
22.	11.1	31.2	31.1	26.6
23.	11.1	31.2	31.1	26.6
24.	11.1	31.2	31.1	26.6
25.	11.1	31.2	31.1	26.6

63 1962-11  
266

## Chapter 2: Part II

### COMPARISON AND ANALYSIS OF RATER RESPONSES TO THE ANCHOR PAPERS IN THE WRITING PROMPT VARIATION STUDY\*

Catharine Keech and Mary Ellen McNelly

#### OVERVIEW OF THE STUDY

Part II of the the writing prompt variation study examines task interpretations of student writers as revealed through their discussion of the qualities of the anchor papers during a training session to prepare them for holistic scoring of their own papers. Comparisons are made of the ways the student interpretations and evaluative criteria differed from those of the expert raters who also scored the papers.

In an attempt<sup>to</sup> identify factors in writing test design which may contribute to the instability of student performance on direct measures of writing, the following report compares student and teacher evaluative criteria in scoring student essays. One of the many variables which must affect test reliability is the unpredicted student interpretation of the writing task. It is sometimes assumed by testers and teachers that students are attempting what they ought to be attempting in response to a given writing prompt. Their writing problems are thought to come from failing to achieve what they set out to do. In this study the assumption is made that sometimes students set out to write something which is different, more difficult, or less appropriate than what the test maker expected them to do. Thus, the finished essay, though successful by the student's criteria, is judged harshly because it does not fit the rater's conception of the task. This lack of congruence in task interpretation between student writers and raters may be a more likely explanation than the assumption that students set out deliberately to write "off topic." Recent research on the writing processes of young children (Scardamalia and Bereiter, 1979) and of adults (Flower and Hayes, 1981) underscores the importance of taking into account the writer's interpretation of the given task, a construction that necessarily includes self-determined criteria for successful completion.

\*THIS REPORT IS BASED UPON RESEARCH SUPPORTED BY THE NATIONAL INSTITUTE OF EDUCATION UNDER GRANT NO. NIE G-30-0034 TO THE BAY AREA WRITING PROJECT, UNIVERSITY OF CALIFORNIA, BERKELEY, LEO RUTH, PROJECT DIRECTOR.

This study then seeks information about student interpretations of writing prompts and evaluative criteria through a potentially rich and generally untapped data source: the peer-rating and the training session for holistic scoring, especially when that training session allows student raters opportunity for full discussion of their criteria evolved in ranking the sample essays in the anchor set. By looking at the evaluations and rankings of student writers when they undertake the role of raters, we learn about their sense of task demands and the values or criteria they hold most important in deciding when the task has been well-completed. By comparing these student rater evaluations to those of expert teacher raters who have evaluated the same student essays, the study uncovers several sources of variation in task construction, evaluative criteria, and the scoring process itself which may affect student scores in uncontrolled or unpredictable ways.

This study addresses two primary questions, each of which includes a subsidiary question:

1. Do writers and raters have the same perception of what the task is?
- 1a. Are some task interpretations likely to generate weaker writing that is generally less competitive in a holistic scoring than writing from other interpretations?
2. Do writers and raters have the same perception of what constitutes a successful completion of the task?
- 2a. Will agreement between writers and raters, on (a) what is the task, and (b) what is a good solution, be better for some kinds of tasks than for others?

(Specifically: will Version B, which specifies audience and purpose in a manner likely to elicit writing for peers, be more likely to create conflicts in interpretations and evaluative criteria than Version A, which elicits writing for the teacher with an unspecified--but implicit rhetorical purpose?)

Following the report, Appendix A provides a detailed discussion of each sample paper used as an anchor during the training of the three groups of raters. The same set of anchors selected by the researcher/trainer, a total of 14 papers, was used for

training all groups. The appendix presents the researcher/trainer's response to each paper, the responses of student raters, the responses of novice teacher raters, and the responses of expert raters. The student responses were recorded and reported by Mary Ellen McNelly, the Bay Area Writing Project teacher-consultant and co-researcher whose students participated in the study as writers and scorers. Responses of the 15 teacher novices and the 4 expert scorers are reported by the researcher/ trainer, Catharine Keech. Responses of the teacher novices (student teachers in the 1981 Writing Credential Program at the University of California, Berkeley) are included for interest as representative of the assessment values and expectations of beginning teachers. The expert scorers are Bay Area Writing Project teacher consultants with long experience in holistic assessment.

## PROCEDURES OF THE STUDY

Because the basic objectives of this whole prompt variation study have been set forth in the Overview under Part I, here the purpose will be restated only briefly: To investigate effects of variation in the specification of audience on a timed-writing test, 114 high school students, grades 10-12, from four classes taught by the same teacher,\* were given a thirty-minute test asking them to write about a particular personal experience, an occasion when the writer experienced something for the first time. Two versions of the prompt were randomly distributed to students in all classes to provide for comparison of different task constraints and their effects on performance. The two versions of the prompt, the student participants, and the conditions of test administration are discussed in Part I of the study.\*\*

### Data Collection and Analysis

After collecting the student essays and a follow-up questionnaire that probed attitudes toward the test-writing experience, the researcher/trainer\* returned to the classrooms and trained the same groups of student writers to participate in a holistic scoring of their essays, using the Educational Testing Service model for holistic scoring, as adapted by the Bay Area Writing Project. The essays were coded and scrambled in such a way that students were unlikely to be able to identify the writer of any essay.

The holistic scoring process was repeated with two groups of teacher raters, referred to in the report as "novice teacher raters" or "novices" (teacher trainees) and "expert teacher raters" or "experts" (experienced teacher raters). Five comparisons were made based on observations of the training and scoring process and on the actual scores assigned by the three groups of raters.

Four statistical comparisons are reported first:

1. Inter-rater reliability estimates. Groups are compared for within-group agreement, or interrater reliability during scoring.

---

\*Mary Ellen McNelly is the teacher/researcher, or "The teacher;" Catharine Keech is the researcher/trainer.  
\*\*The test instrument and instructions to students appear in Appendix A of Part I.

2. Agreement among groups. Scores given to individual papers by all three groups of raters are correlated as a measure of between-group agreement in rating the essays.
3. Differences in scoring parameters among rater groups. Means, standard deviations, and frequency distributions of the scores given by each group of raters are compared as indices of each group's general impression of the whole sample of essays as well as each group's use of different points in the scoring range.
4. Differences in score distributions for three rater groups. Differences in the groups' responses to the two different prompt versions are reported, as measured by mean scores and standard deviations of scores assigned.

Finally, a qualitative comparison of the rater differences in responses to the anchor papers during the training sessions is made. This comparative description which allows further interpretation of the statistical findings suggests differences in writer and rater attitudes which may not be clearly reflected in the statistical data.

#### Raters As Subjects

It should be noted that in this second part of the study, the research subjects are not the students-as-writers, (as they were in the first part); rather, the subjects now are the students-as-raters and two other groups of raters who all evaluate the student writing on the two "first time" prompt variations. There are then three distinct groups of raters as subjects in the study, in addition to the researcher who also becomes a subject of analysis in her role as trainer of all groups of raters for the holistic scorings.

1. Student Raters. The students involved in the holistic scoring of essays were drawn from the four classes who were administered the writing test. (The make-up of this population is fully described in Part I of this study.) It should be noted that only three of the four classes tested were used for scoring the papers, and not every student within each class made a commitment to a particular score during the initial tallying of responses. The sophomore non-accelerated class was excluded from scoring because it was considered to be potentially less accurate and consistent. In addition, the number of student responses as recorded below varies from 27 to 84. The low counts are derived from the second scoring session when only

one class was retrained using the remaining anchors. The high counts were taken in the first scoring sessions, held in all three classes. Not all students voted on every paper during training.

## 2. Novice Teacher-Raters

In a second scoring of the essays, teacher trainees in the Writing Teacher Credential Program at the University of California at Berkeley were taught holistic scoring using the student essays produced for this study. As with the student raters, the numbers of responses for each anchor paper vary. Twenty-three teacher trainees participated in the first tallying of responses to Samples A, B, C, H, I, J, and N, with only 15 teacher trainees participating in the second tallying using the remaining samples. These 15 novice teacher raters also did the actual scoring of the student essays.

## 3. Expert Teacher Raters

Four Bay Area Writing Project teacher consultants were hired. All had long experience in holistic scoring for their own school districts as well as for the Educational Testing Service. All had scored the California High School Proficiency examination, and several had read the Advanced Placement examination in English and/or the English Composition Test given as part of the College Board examination. Designated as expert raters, all four had taught high school English for 20 years or more each, to students similar to the those in the sample.

### Anchor Papers

The "anchor papers" are selected in order to allow the establishment of scoring categories that can be easily distinguishable by raters. They represent papers which combine both strengths and weaknesses in ways typical of other papers in the sample. Thus, examining and discussing these anchor papers provides readers with models for making accurate and reliable rapid impression decisions while assigning scores to the remaining papers.

The trainer works within the limits of the scoring points allowed: for this study, a range of 1-4 was used, with 1 representing the poorest papers and 4 the best papers in the set.

Typically a trainer of holistic raters (who may be assisted by table leaders) reads from 10 to 20 percent of all the essays in order to select from 10 to 30

sample papers which can be used to "anchor" raters during training. Common experience suggests that the larger the number of papers read before choosing anchors, the greater the likelihood of representing in accurate proportions, the entire range of writing quality in the total sample, and the less the likelihood of selecting a paper which is atypically high or low. Such a paper would make a poor anchor because it would establish too narrow a range of criteria for a particular scoring point.

In the current study, the researcher/trainer read the entire sample of 114 essays before selecting anchors in order to gain a sense of the range of quality as well as a sense of what would constitute upper-half versus lower-half papers. Besides including best and worst essays and a representative selection of middle range papers, the trainer selected one paper which seemed to constitute an evasion, or refusal to write about the assigned prompt (Sample M, discussed below), and one unfinished paper which broke off in mid-word (sample F, below). These latter two provided examples of typical scoring problems. Papers representative of responses to both prompt versions A and B were also chosen. In addition, papers in several forms, such as diary and letter, as well as informal essays (the most common form used), were chosen to allow raters to discuss any scoring problems that might arise from attempting to compare quality of writing across genres.

### Training the raters

In training the student raters and the novice teacher raters to rank order a set of sample papers, or anchors, the researcher/trainer observed which papers the raters found superior and elicited from them the reasons for their choices. Thus, their implicit standards for writing this particular kind of paper, their interpretation of the writing task, as well as other constraints surfaced as they evaluated and ranked the anchor set.

### Training the Student Raters

After a brief introduction to the nature and purpose of holistic scoring, sets of anchor papers were distributed to all students in one class period. Students were asked to read quickly through anchors A, B, C, H, I, J, and N, forming an impression of the range of quality in this set. They were told to choose the paper they thought most outstanding and give it a score of 4, after which they should choose the poorest paper and give it a score of 1. They were told that there might be more than one paper they thought was best, in

which case they could give two scores of .4. In the same way, they were allowed to give two scores of 1. After establishing the outer boundaries of the scoring range, they were directed to rank order the remaining papers until all papers had scores and all scores had been used at least once.

The student raters were trained during their regular English class periods. The training sessions were consistent in all three of the upper-level classes used for the scoring. The votes generated for a particular paper fell into similar patterns in all classes. It proved possible for the researcher/trainer to maintain a consistent scoring categories across all of the class periods.

As students read the anchors, the researcher/trainer placed a grid on the board as shown in Figure 1.\* Then when they were ready, the researcher called for a show of hands to indicate which scores students had assigned each paper. The tally revealed that the majority of students ranked Example A exactly as the researcher-trainer had, although some students had ranked it higher or lower, as is usually the case in the early stages of training inexperienced raters. Discussion of example A followed in which low raters cited its weak points, while high raters defended their higher ranking. This discussion is the heart of the training process, allowing raters to become sensitive to qualities and flaws that they had not previously attended to, as well as helping them to adjust the relative importance they attach to different strengths and weaknesses in the light of how successful the piece of writing is on the whole, compared to other writing in the total sample. They begin to develop a sense of what a 2 versus a 3 versus a 4 might mean for this sample of papers. It is also during this process that the students revealed what they valued in a piece of writing: what their criteria for good writing were and how they had construed the task to begin with. (Their evaluative responses to each anchor paper are summarized in Appendix A, along with the responses of the teacher raters.)

As training progressed within each class period, student raters in each class began to agree more closely on their rankings of papers which were clearly representative of a particular scoring point. One or two papers, however, were awarded a broad spread of scores; these would typically cause discrepant scoring

---

\*All figures and tables are placed at the end of the report.

if encountered during the reading. As useful as sustained analysis of such papers is, there is also a danger in analyzing anchors too closely. Over-extended discussion of specific features of a paper may mislead raters, costing them their ability to form a whole impression of a work and making it difficult to compare it to other papers for the overall effectiveness of its unique combination of flaws and strengths.

In the training of the students raters, as in the training of any group of inexperienced readers, the rather open discussions of each paper were concluded with a strong summary of its general worth to confirm the majority vote and to clearly designate the essay as belonging to one scoring category or another. Whenever an essay appeared to fall between categories, a split vote was sometimes allowed to stand, since prolonged discussion or arbitrary rulings often serve to make the raters insecure about their category boundaries. In such cases, students were reminded that each paper always receives two scores, and that the combined ratings on such a paper would be likely to reflect the two views, yielding a final score (of, for instance,  $2 + 3 = 5$ ) which would be descriptive and fair.

#### Training the Novice Teacher Raters

The training of an initial group of 23 novice teacher raters was part of a general workshop on holistic scoring and followed an introduction to assessment theory not given to the student raters. However, in all other respects, the same training procedures described above were used, with the researcher/trainer eliciting the novice teachers' untutored responses to each anchor paper before attempting to summarize and achieve a consensus. This first group of 23 student teachers tended to use the training session more as a forum for exploring and shaping their own criteria for evaluating writing. In general, classroom experience had not yet provided them the opportunity for testing their ideas the classroom. Consequently this novice teacher group's operative criteria, as expressed during the training, never stabilized during the first training session to the same extent as the student raters' responses had done. A second training session was held a week later with 15 teacher novices from the original group who proceeded to score the entire sample.

#### Training the Expert Teacher Raters

The same set of anchors used to train student raters and novice teacher raters was given to the four expert teachers for ranking. They quickly reached a high rate of agreement. However the speed and

confidence with which they reached agreement on the proper scores to assign each anchor paper proved to be misleading and led to a problem in defining certain categories during the reading. Even though the researcher/trainer observed that the expert raters seemed to be defining the category of 3 too broadly and feared that too few papers in the full sample would fall into the 2 category, she honored the distinctions that the experts wished to draw, believing that letting them determine the scoring boundaries, which seemed so clear to them, would increase the reliability of their scoring.

The researcher/trainer chose not to intervene to narrow the 3 categories also because she was aware that the larger proportion of students in the sample were quite good writers, with only one class in four made up of younger, less experienced writers. She thought the experts' classification of the papers might more accurately reflect the actual distribution of ability in this skewed group. And the researcher/trainer recognized that it is an innate problem of holistic scoring and other rank-ordering that the original scale which is applied to a set of papers does not necessarily reflect equal distances between integers. There may be inherent in setting the scale a conflict between the practice of trying to distribute students in a fairly normal distribution across scoring points and the practice of trying to arrive at scoring points that are equal intervals apart in terms of the quality of essays, rather than in terms of the number or percentage of essays that fall in each interval. Having earlier defined the intervals by trying to distribute students across all scoring categories in something approaching a normal curve (albeit skewed by nature of the population), the researcher-trainer withdrew from the discussion of the experts, who seemed to be drawing on a common experience and making reference to outside criteria to determine the scale point intervals in terms of quality of writing.

As the reading progressed, however, it became clear that the expert themselves believed that they had defined the category of 3 too broadly in scoring anchors. They described seeing the 3 category as "high threes" and "low threes"--a common practice in large readings, where a few scoring points are used to reflect a broad range of writing quality. But they noticed that other categories such as 2 or 4 did not get broken up in their minds. They agreed with one of the raters who said that a five point scale was needed for these papers to deal with their feeling that too large a range of papers received a score of 3. The papers with this score were more different from one

another than were the papers within any other scoring interval.

### Rating the Papers

Student raters spent most of the first day receiving training in holistic assessment, using the anchor papers. In the final 10 minutes of each 50 minute class period, papers were distributed randomly to the class, two to each student. Essays were coded, and essays from all classes were included in the sample distributed to any one class. (Students were instructed to exchange papers with another student if they had their own or thought they recognized the writer of the paper they were scoring.) Students were informed that the teacher and the researcher/trainer would be spot checking the holistic scores assigned in the interest of fairness to the writers, but that in most cases, the total score assigned by the student raters would constitute the writer's score on the essay. Students were again reminded that their rating was only one of two the paper would receive.

The scores given on the first reading were concealed before the student raters passed the papers along for the second reading.

On the second rating day, students in the two largest classes (classes which had shown the greatest stability during training) completed scoring the essay set after being retrained with the remaining anchors, samples D, E, F, G, K, L, and M. Results of their discussion of the latter anchor set are also included in the appendix. The teacher and the researcher/trainer re-scored one third of the essays and each found only one paper which was inaccurately scored by both readers. Interrater reliability estimates are reported in Table 1.\*

Novice teacher raters began and completed scoring in one two-hour session, after being retrained with the same set of anchors used in the student retraining. The researcher/trainer acted as table leader, moving around the room to check the ratings of each rater at least twice during the two hours. Four tables of readers drew from randomized piles of essays in the center of the tables and concealed the first ratings from the second raters. Papers scored at one table were distributed to another for the second ratings. After the scoring had ended, all discrepantly scored papers were given a third reading and a new total

\*All tables appear at the end of this report.

score. Inter-rater reliability estimates for this group are reported in Table 1.

The four expert teacher raters readers sat with the researcher/trainer around a single table for the training and scoring of the sample. Papers were randomized and divided into eight stacks. Raters exchanged papers in a rather complex arrangement intended to insure that every rater was matched to every other rater for roughly equivalent proportions of the reading. All discrepancies were resolved through provision of a third reading during the same rating period. Inter-rater reliability estimates for this group are reported in Table 1.

287

## STATISTICAL COMPARISON OF RESULTS FROM THREE GROUPS OF RATERS

There are a number of ways to describe the amount of agreement among raters, both within the three rating groups, and among the groups in a comparison of the three ratings of the same set of student essays. The first two statistics of importance measure within-group rater agreement, or what is usually known as inter-rater reliability.

### Inter-rater Reliability Estimates

There are two accepted ways of describing the amount of agreement between raters of writing samples in holistic assessment. The first involves simply reporting the number or percentage of essays which received scores from two raters with more than a one point difference. With each rater using a scale of 1 to 4, so-called discrepant ratings would include, for example, a rating of 2 from the first rater and a rating of 4 from the second rater, or paired ratings of 1 and 3, or 1 and 4.

Reported in the final column of Table 1 are the discrepancy percentages, or percentages of total essays which received discrepant ratings from two readers, entered as  $d = .078$ , for example, or 7.8%. As training procedures have improved and the pool of experienced readers has increased, local holistic assessments in California typically report discrepancy percentages of less than 10%, many regularly achieving less than 5%, and some places, such as San Francisco State University, which maintains similar criteria from year to year, report consistently achieving fewer than 1% discrepancies.

When the students who wrote the essays assumed the role of raters and were trained in holistic scoring, they achieved a respectable 7.8% discrepancy rate, while novice teacher raters, who proved less responsive to training, produced a 15% discrepancy rate. As expected, the experienced teacher raters maintained a low, (4.4%), discrepancy rate.

A second estimate of inter-rater reliability is usually reported as the degree of association or the correlation between the scores of the first and second raters on all papers. The measure used here is Pearson's  $r$ , with perfect correlation represented by 1.00 (all first and second readers agreeing exactly on the scores assigned to each paper), and no correlation represented as .000 (no first and second readers agreeing exactly on the score of any paper, with both

maximum amount and random direction of disagreement; however, this is a theoretical, not a practical, possibility). Values for Pearson's  $r$  of .01 to .35 are considered quite low, for most purposes; correlations better than  $r = .35$  up to but less than .65 are generally considered moderate rates of agreement, while correlations of more than .65 are considered fairly strong rates of agreement. In the early days of holistic assessment, the best rates of agreement between two readers were rarely over .60 (Diederich, 1974). More recently, testing programs, including not only the professionals at Educational Testing Service but many local schools, have been reporting raw, unadjusted correlations between readers of .7, .8, and in some instances of better than .9.\*

Table 1 reports the correlations or inter-rater reliability estimates for the three groups of raters, showing the student raters achieving a statistically significant correlation of  $r = .68$  ( $p = .01$ ) which approaches a generally acceptable rate of agreement by today's standards. The novice teacher raters, on the other hand, produced only a moderate correlation of  $r = .55$  ( $p = .01$ ), which is considered a poor rate of agreement for raters. The experts achieved a correlation of  $r = .71$  ( $p = .01$ ), as good a correlation as that achieved by experienced raters in many local assessments, and in some professional assessments, as well, but lower than anticipated by the researcher-trainer who expected the four experts chosen to be highly compatible readers. The discussion below considers possible reasons for this less than outstanding rate of agreement.

#### Agreement Between Groups

For comparing scores assigned to the same essay by the three different groups of raters, only the essays' final or total corrected holistic scores were used (ranging from 2 - 8 possible points). For instance, an

---

\*Diederich reports a procedure for "stepping up" reliability estimates by using two or more samples of student writing. This cannot be done, of course, where only one sample of writing is rated for each student. Also, since discrepant scores are always resolved by giving the paper a third rating, which is used to provide a "corrected" or adjusted score for the essay, it is possible to calculate the measure of association using these corrected scores. This is not generally done, however, since testers usually wish to estimate the extent of agreement between the original two raters, to test the accuracy of the reading as a whole.

essay might have received a score of 5 from the student's rating, (representing the two scores of 2 + 3), while the novice teacher raters rewarded the same paper with a total score of 6 (3 + 3), and the expert teacher raters judged it even more favorably, giving it a score of 7 (3 + 4). Pearson's  $r$  is used again to report correlations between the total holistic scores awarded by each group.

Table 2\* reveals that the poorest agreement was between student raters and novice teacher raters at  $r = .47$  ( $p = .01$ ). The best agreement was between novice teacher raters and expert teacher raters, at  $r = .62$  ( $p = .01$ ). Students agreed with the expert teacher raters only moderately,  $r = .50$  ( $p = .01$ ). (Novice teacher raters did not score 8 of the papers because duplication of original was inadequate for rating. Student raters and expert teacher raters scored the full sample, working with original papers.)

These simple correlations merely alert us to differences between the groups of raters in valuing the papers; the following statistical analysis, together with the narrative descriptions of rater training and responses to specific essays, provide more insight into the nature and extent of those differences.

#### Differences In Scoring Parameters Between Rater Groups

Table 1 reports means, modes, medians, and standard deviations of the scores given to the full set of essays by each group of raters, respectively. Thus, we can see that, if the student's own ratings were used to estimate the writing abilities of the groups who wrote this set of essays, the entire group would earn a mean score of 5.66, with a score of 5 (3 + 2) being the most common score awarded (the mode). Novice teacher raters provide a more positive estimate of group ability, the average of all scores awarded being 5.95, with the most common score being 6. Expert teacher raters agree with novice teacher raters in producing more ratings of 6 than any other score, but reward the set of essays with an even higher mean score of 6.18.

The difference between these means may not appear to be great, but comparing expert teacher raters to student raters (the original comparison of interest), the difference is .52, about half a scoring point (statistically significant,  $p = .05$ ) or just over 1/3 of a standard deviation. Such a difference between holistic scores, earned by different groups of writers, is

\*Tables appear at the end of this report.

usually considered educationally significant, representing a greater difference than usually occurs when two groups of similar ability are compared after writing on the same essay topic (prompt), or when the same group is compared before and after instruction, writing either on the same or on different essay topics (prompts). In other words, student evaluations of the papers in this set are significantly lower than teacher evaluations of the same set.

### Differences in Score Distribution for Three Rater Groups

Reporting only the numerical parameters of the scores assigned by each group, however, obscures additional important differences in the scoring patterns of these groups. Such differences invite interesting speculation, although they generally support the finding that these students seemed willing to judge their work more harshly than experienced teacher raters did.

Table 2.2 shows the distribution of scores given by the three rater groups. It is evident at a glance that the student raters came closest to providing a normal distribution of scores of the sort expected in large-scale holistic scorings. They were more willing to use the lower end of the scoring range, with six papers receiving a score of 1 from at least one reader (total scores for these papers would be either 2 or 3, depending on the second rating). Novice teacher raters used fewer scores of 1 (on only three papers), and the expert teacher scorers gave only two papers a 1, and even then, these papers both received a score of 2 from their second rater, so that not one paper in the sample was given the minimum score of  $1 + 1 = 2$ . This occurred despite the fact that the expert teacher raters had agreed to give a 1 rating to one of the anchor papers thereby indicating it as typical of the category of writing. When encountering the same paper during the rating, at least one expert teacher rater changed her mind and gave it a 2.

Student raters also had a broader sense of the 2 category of paper, distributing papers in the middle range (scores of 2 or 3) equally across both score categories, with the same number of papers (23) receiving two ratings of 2 (total score = 4) and two ratings of 3 (total score = 6). Slightly more papers (26) fell halfway between with scores of  $2 + 3 = 5$ . Novice teacher raters, on the other hand, awarded a score of 2 to fewer papers, preferring to use a 3 score for a broader range of essays. The expert teacher raters carried this tendency even further and placed 47 papers in the  $3 + 3 = 6$  category, with only 10 papers

receiving scores of 5 and only 13 receiving scores of 4. Possible reasons for this odd distribution are suggested in part by differences in the training of the groups, and are more fully explored below.

### Discussion of Results

The differences between groups in the inter-rater reliability achieved are perhaps not surprising. The students were systematically trained; the criteria for good and poor essays, after being elicited from students, were re-articulated and delineated by the researcher/trainer, and the students appeared eager to cooperate, to "learn how to do it right." As a result, they achieved a quite respectable correlation between raters of .68. With the novice teacher raters, on the other hand, there was less willingness to reach consensus. They persevered with their idiosyncratic readings of various essays, exactly as one might expect a group of inexperienced teachers to perform without benefit of moderating influences accumulated through classroom experience. The result was a low correlation between novice teacher raters of only .55. The expert teacher raters were confident and compatible in their opinions, having done many scorings together for the Educational Testing Service. No attempt was made to moderate their views, with the result that the scores they gave to the anchor papers may have been less discriminating guides than they might have been. But even so the expert raters achieved a consensus which in turn resulted in a reasonably good correlation between first and second readers of .71.

Failure to achieve a closer correlation among the expert readers may be attributed, at least in part, to certain factors which may affect reliability in any holistic scoring. The two versions of the prompt elicited, in some cases, very different content and modes of writing, with the audience-emphasis version evoking from some students what might be called process essays, or directions on how to do a particular thing for the first time (See Sample D, Appendix A). These latter essays were different in many respects from the typical response to either version of the prompt, which usually was narrative in structure, and provided an anecdotal account of the writer's own "first time." Other work in this assessment project (See Technical Report No. 5) illuminates the difficulty that raters have in comparing two very different types of writing in a single rank-ordering based on over-all quality. In general, as shown in the qualitative description of the training sessions (See Appendix), the process papers suffered by comparison with the narrative, autobiographical papers; however, some raters responded more positively to the

process papers, believing that giving directions is a more difficult task than narrating an event. Thus they over-rewarded the writers of these papers.

Another factor which may have contributed to the disagreement among raters is the instruction on the writing test: "You may write an essay, short story, letter, journal, etc." It is difficult to make fair comparisons among pieces written in these different genres. In addition, as can be seen in the discussion of responses (Appendix A) letters and diaries present special problems to readers who may choose either to forgive or not to forgive infelicities of expression, informal language, casual organization, and lack of development or lack of vivid concrete details, when these features represent the student's accurate interpretation of the tone and style of a real letter or diary entry.

Some of the disagreement revealed in the low correlation of scores between groups and in the striking differences in scoring distributions may be attributed to the training and the initial defining of the scoring categories. During training, the student raters were encouraged to consider virtues and faults of the anchor papers identified by the researcher/trainer. More amenable to the training, they were more likely to divide the sample in a way that distributed papers normally across scoring intervals. On the other hand, the expert teacher raters who rapidly defined their scoring categories may have drawn on prior definitions of the scoring points. Possibly they may have been influenced by the five-point scale they were each familiar with from scoring the California High School Proficiency Exam (CHSPE) and by the quality of essays drawn from the general high school population that they typically encounter while scoring that test.

More directive procedure for training the expert teacher raters might have yielded different results. Or if the expert teacher raters themselves had read the entire sample before pulling anchors, rather than having the researcher/trainer choose the anchors, the distribution of scores might have more closely resembled the students' distribution.

There is no question that the process of choosing anchors to define a scoring range is a critical step in any holistic assessment. Choosing anchors requires the juggling of what may be contradictory constraints. On the one hand, there is the need to distribute the sample quantitatively according to some criteria--either that the distribution represent the distribution of the

population on other measures of related ability, or that the population be equally distributed across scoring points, or that the population be normally distributed across the scoring range. On the other hand, the further constraint--that the scoring intervals should represent roughly equivalent differences in quality among essays--is also imposed. In practice, the latter constraint may be defined by the former: testers may come to feel that the operational definition of "equivalent differences in quality" from one scoring point to another should be that "equivalent or appropriate numbers of essays" are grouped into each scoring interval. This complex interplay of quantitative and qualitative considerations, an interplay which takes place subjectively for the most part during the selecting of anchor papers, deserves further study.

Even if student raters and teacher raters arrived at similar distributions, with similar mean scores, the agreement of their verdicts on particular papers, represented by the correlation between student rater and teacher rater scores, might still remain low, for the three rating groups worked within three distinctly different frames of reference for judging the essays. The student raters' expectations very likely were derived from familiarity with their teacher's expectations, with their own and their classmates' work, and with the standards generally set forth in their particular high school.

The expert teacher raters, none of whom teach at the school where the study was undertaken, had expectations drawn from a much larger assessment population, including not only their own schools and classes, but also the large number of students throughout California who write for the proficiency exam which these teachers have scored since its inception. In general, the student writers in this study produced better papers than might be found in these larger school and state populations. Consequently by this standard, the expert raters found fewer really poor papers.

Finally, the reference points for the novice teacher raters became somewhat confused, and certainly mixed. The between-group correlations in Table 2 show this latter group to be closer in agreement with the expert teacher raters than they are among themselves. This condition is easy to understand if we imagine that two novice teacher raters might produce a discrepant scoring of 2 + 4 on a paper which two expert teacher raters might recognize clearly as a 3 + 3. Both novice raters would be closer to the experts than they were to each other, which would explain the pattern of correlations reported.

Additional explanations for the lack of agreement among the three groups of raters are best sought in a closer examination of the values and criteria expressed during the training sessions, in reaction to particular anchor papers. The Appendix includes individual and group responses recorded during training. It also reveals the criteria which seemed most operative for the three groups of raters as well as for the researcher/trainer.

One interesting within-group phenomenon remains to be explained. Why did the expert teacher raters agree to assign the score of 1 to at least one anchor, but fail during the actual reading to give that same paper a score of 1 on both of its readings? The most likely explanation is that these teachers were working with pre-established notions about the kind of writing which should fall into the lowest scoring category. In many local high school assessments, as in the CHSPE, scores of 1 are reserved for papers with so little writing as to be nearly unscorable, or with such severe scribal or transcription difficulties as to be nearly indecipherable and severely flawed in spelling and punctuation. To keep standards constant from one year to the next, these criteria are carried over, even as students taking the test produce fewer and fewer papers in this lowest category. Thus, the expert teacher raters may have held to an external criterion for this lowest score, even though it meant they almost never used the scoring category. The other groups, without this background experience, were more willing to follow the directions and try to use the full point range. As is evident from anchor paper (B), which was scored a 1 during training, evaluative criteria drawn from other general readings would be inappropriate for this sample.

There is one more factor which may have affected scoring results in unaccountable ways: In the reading and rating of the anchor during the training session typed versions were used. But during the actual reading, these same papers were in their original handwritten forms. The expert teacher raters reported that they had different feelings about the anchor papers on reading them in their original handwritten form. Unexpectedly, they tended to want to give the handwritten versions higher scores, in spite of messy handwriting or less than attractive presentation. The typed versions apparently made these teachers more conscious of slips of the pen--words omitted or repeated, faulty capitalization or spelling that could have been errors in transcription rather than errors of student knowledge or performance. The errors appeared more glaring when found on the printed page, than in the

context of an impromptu hastily handwritten essay, where they were generally ignored as being indicative of little more than lack of time to proofread.

This is an interesting finding, in the light of previous work (Redmondino, cited in Diederich, 1974, p. 9) which demonstrated that handwriting is one of the factors affecting teachers' judgments of an essay, with poor handwritings lowering scores, and good, clear handwriting positively influencing the reader. In this study, however, a clear handwriting might well have the effect of making errors more visible. It is possible that the student who blurs letters in order to render possible misspelling less obvious may be using an effective strategy-- at least in hastily written and holistically scored timed essay tests.

## CONCLUSION

A major significant finding of Part II, is that some students do in fact interpret the writing task in ways which differ from teacher-rater task interpretations. Yet these variant interpretations are reasonable, considering the instructions in the writing assignment. The greatest mismatch between teacher and student expectations and evaluative criteria appears to occur when students are assigned the prompt versions which encourages them to write to a specified audience other than a general or teacher audience. For example, when students write letters to friends or parents, or make entries in personal diaries, many of them value verisimilitude in such instances, and they attempt to make such pieces sound more like "real letters," than like "real English themes." The conflicting frames of reference provided by the simulated life-like communicative situations and the actual writing context, a classroom test occasion, can be successfully integrated only by the most able students. Even then, that success may be questioned by student raters whose understanding the nature of real letters and diaries does not allow them to accept the intrusion of any hint of school writing, which is seen as unnatural or showing-off. Consequently, there are disparities between the pupils' conception and execution of assessment tasks and the expectations of those who author the prompts and rate the performance in response to them.

TABLE 1  
SCORING DISTRIBUTIONS FOR EACH GROUP OF RATERS

		FULL SAMPLE	VERSION A	VERSION B	RELIABILITY
STUDENT RATERS	X	5.66	5.65	5.67	d = .078
	$\bar{m}$	5.00	4.00	5.00	r = .63*
	M	5.53	5.55	5.62	N = 109
	S.D.	1.49	1.64	1.33	
	n	114	57	57	
NOVICE- TEACHER RATERS	X	5.95	5.67	6.26	d = .150
	$\bar{m}$	5.00	6.00	6.00	r = .55*
	M	6.01	5.78	6.19	N = 106
	S.D.	1.33	1.40	1.20	
	n	106	55	51	
EXPERT- RATERS	X	6.18	6.09	6.23	d = .044
	$\bar{m}$	6.00	6.00	6.00	r = .71*
	M	6.13	6.10	6.29	N = 114
	S.D.	1.27	1.29	1.25	
	n	114	57	57	

d = discrepancy rating: proportion of total essays requiring a third reading.

r = Pearson's r.

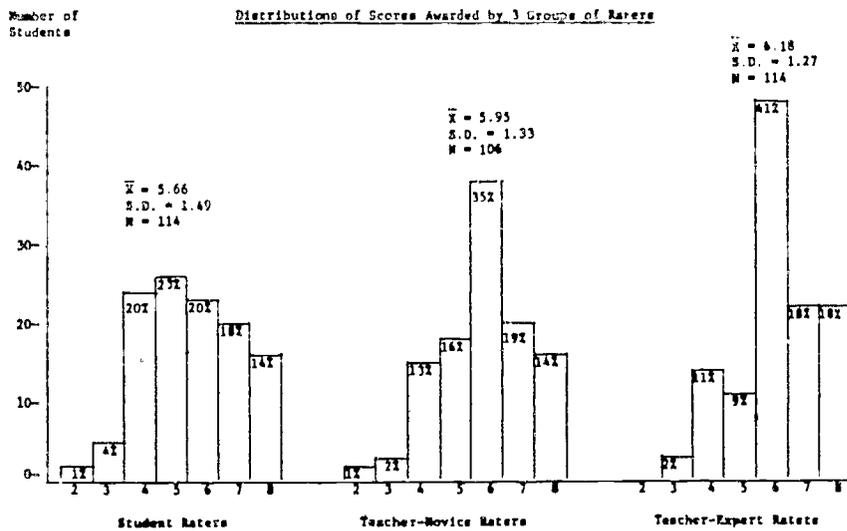
N = total papers used in calculating d and r.

\* = statistically significant,  $p < .01$ .

TABLE 2  
AGREEMENT BETWEEN GROUPS OF RATERS

Rating Group's Scoring Same Samples:	Pearson's R =	Significant	N
Student raters with novice teacher raters	.47	.001	106
Student raters with expert teacher raters	.50	.001	114
Novice teacher with expert teacher raters	.62	.001	106

Table 2.1



301

Figure 1

Facsimile of Grid for Period Five: 32 Students

	A	B	C	D	K	L	anchor paper
1		23	8		2	3	
2	8	9	20		7	19	
3	16		3	10	11	8	
4	6			22	11	1	
Total Votes:	30*	32	31*	32	31*	31*	

\* Not all students voted on every anchor.

30<sub>12</sub>

## APPENDIX A

### COMPARISON OF SCORES GIVEN DURING TRAINING BY THREE GROUPS OF RATERS AND THE RESEARCHER-TRAINER

Each of the student papers used as an anchor for training the three groups of raters is transcribed below in full, exactly as it appeared to the raters. Following each essay is a summary chart showing distribution of rankings given by each group for this particular essay. These rankings are followed by a qualitative description of rater responses to the essay, including the researcher/trainer's evaluation of the paper's strengths and weaknesses which influenced her choice of the paper to represent a particular score during training.

304

APPENDIX A

SAMPLE A

As I look back on my attempts at new and different things, one experience that sticks out and under the rest is futile effort at driving a car with a manual transmission. "The Clutch," the dreaded part of the car for beginners, causes more fits of flustered frustration than any other mechanical device. It seems to sense a first-timer and tends to be a little bit tighter at first as compared to when you get familiar with it. I remember trying to stall the car one time at a stop sign just to prove to me that it was the "Clutches" fault for all of those mid-intersection stalls it pulled on me. I know they try to make your engine die when your trying to impress the girl you like whose waiting at the crosswalk, and they try to make you lay a little layer of rubber on the road when a cop is in close vicinity. Everybody knows all of the tricks a clutch pulls.

Unless you like the harmony of ten different car horns, don't don't try a main intersection on your first day on the "clutch." Its been over a year now that my mom happened to bring me up form a skate-boarder to a "Driver." We'll set aside the countless of times I've popped the clutch or stalled the damn thing. None were more memorable than the stunt I pulled at the Blossom Hill--Los Gatos Blvd. intersection. I was the first in line at the left turn Signal with the simple thought of making a U-turn when the light turned green. I was easier that done back then, and as the light turned green, my sweaty hands shoved the car in gear and the jittery feet didn't quite put the "ole gas and clutch" combo together right and the car the preceded to advance nowhere. There was a non-melodious choir of horns to follow and I then attempt to go again, only to do an encore, then finally, as the light turned yellow sacrificed the clutch and burned out of the lane and left 20, angry motorist.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	0	(0)	1	(04)	0	0	0
2	8	(10)	5	(22)	0	0	0
3	51	(63)	13	(57)	3	(75)	x
4	22	(27)	4	(17)	1	(25)	0
CONSENSUS SCORE:	3		3		3		3

Discussion of Sample A

Researcher/Trainer Response. This paper seemed typical of many essays in the sample which might be appropriately rated 3 because of its particular combination of flaws and strengths. This is the voice of a confident writer who is clearly enjoying the assignment, attempting

humor with some success, capturing the frustrations of a familiar experience. There is superficial carelessness, but there is also ample evidence of good linguistic control so that dropped letters or misspellings might fairly be regarded as "slips of the pen" -- certainly they do not interfere with communication. Even so, the number of purely mechanical faults would, in most assessments, bar this paper from a top score. In addition to cosmetic flaws, organization is loose, a result of a very nice attempt to balance first-time troubles against all the ongoing troubles a car with a clutch can give. The "first time" is really, in this writer's mind, a series of experiences, summed up as "trying to drive a car with manual transmission." As if he senses this, he finally focuses in the second paragraph on a single incident: "None were more memorable than the stunt I pulled at the Blossom Hill ...intersection..." Actually, he may have had this episode in mind all along, but he has set it among several generalities about the war with the clutch that all inexperienced drivers must face -- a perfectly good rhetorical device for solving the given problem.

Student Rater Response. The students were spread over three scores in their evaluation of this paper -- no one thought it was the poorest paper. Those who gave it 2's were reacting almost exclusively to their feeling that the writer did not address the topic properly by writing a clear account of a first time experience. One commented, "I couldn't tell what he was really writing about." More favorably impressed students argued that of course one could tell what he was writing about: driving a car with a clutch. And the majority agreed that he wrote about it well -- they liked the writer's humor and details. It is clear in retrospect that for all groups of raters, including the researcher/trainer, this paper might have received the highest score were it not for the number of mechanical flaws and were it read in comparison to any other papers except Sample J.

Expert Teacher Rater Response. One expert argued well for the strengths of this paper -- the other three concurred that the mechanical flaws should prevent it from receiving the highest score.

Novice Teacher Rater Response. Teacher novices were spread over all four scores, as might be expected given the variety of their backgrounds and training in English together with their lack of experience in the classroom. The very low scorers were responding almost entirely to the mechanical errors in this paper, which provided an opportunity to explain the role of error in holistic assessment, to remind them that errors accompanying hasty writing

on a timed test but that do not interfere with communication may be evaluated differently from errors on a carefully edited paper, or errors which make comprehension difficult. The teacher novices who liked this paper argued for it well, with several pointing out that humorous writing is difficult and that they had very much enjoyed the voice and personality of the writer.

SAMPLE B

The first time I ever got drunk was back in eighth grade. The reason that it is memorable for me is because it was such a weird feeling. Me and a couple of friends went up to a party that this girl that we knew was having. She had a couple kegs of beer. I had one, and thought that it wasn't all that bad, but nothing was different. I had a couple more, and I started to feel a little bit strange, but really thought nothing about it. I really did not think that I got drunk, until the next day when a couple people I talked to were telling me some of the things that I was doing, I didn't even remember. This has not become a habit, I get drunk maybe every two weeks or so. Now I know how I'm going to feel, so that is why I can remember the first time that I got drunk.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	61	(73)	17	(73)	0	0	X
2	23	(27)	5	(20)	4	(100)	0
3	0	0	1	(07)	0	0	0
4	0	0	0	0	0	0	0
CONSENSUS SCORE:	1		1		2		1

Discussion of Sample B

Researcher/Trainer Response. This paper was my anchor choice for the lowest score. Among papers in the sample, it was one of the least compelling on a quick reading. The vagueness of the expression, "it is memorable because it was such a weird feeling" is never relieved with the kind of clear description that would show us exactly what the writer really experienced. As it turns out, what is most memorable about the experience is that the writer couldn't remember the experience. Often, the writer may set himself a difficult task, as in this case, in which the choice of episode presented a difficult technical problem: what was interesting to the writer--the fact that he couldn't remember anything--is hard to write about interestingly for the reader. In spite of these disclaimers, the paper was boring to read and had few if any redeeming features. The language is fairly flat and empty, the organization loose and undeveloped.

Student Rater Response. Student scorers easily identified this paper as one of the poorest. However,

its defenders were vocal, and its defense consisted almost entirely of such comments as, "I really know how that feels." One student who voted for a score of "2" waxed eloquent about what happens the first time one gets drunk, and all the things you learn afterwards about what you did. This oral presentation of the subject provided a fine example of what this paper needed to have been in order to receive a higher score, and it helped students recognize the difference between a paper they liked simply because it is about a familiar experience and a paper that presents that experience vividly.

Expert Teacher Rater Response. Surprisingly, the experts chose only one paper among the anchors to give a score of "1" and it was not this one. This paper they compared to Sample C in quality, rating both papers "2". Having encountered 100% agreement, and following ETS training conventions, I did not insist on a great deal of discussion, allowing Samples B and C to stand as representative of the second level score. I was then no longer so convinced that they did represent different rating levels, as I had been by my initial impression of the sample. However, after the expert reading, as I became increasingly aware of how similar in quality were the papers ranked 1 and 2, how few 1 scores were being awarded by raters (because of that similarity), and how large the range of quality was for papers awarded a score of 3, I realized that a better division of scale levels might have ranked Samples B, C, and I at level 1, forcing readers to make a different distribution of 2 and 3 level scores.

Novice Teacher Rater Response. This group was distributed very much like the student raters, most agreeing that this was one of the poorest papers in the sample, but a few saying merely that they "liked" it, that it described a common experience accurately. The one high rater recognized the potential cleverness in taking a subject where the point is that nothing was remembered.

SAMPLE C

Tues., dec. 3

Dear Diary,

This may sound stupid, but I gave a speech today and it scared me half out of my wits. I have given speeches before but this time I could only use notes instead of reading it. I was only allowed five note cards and it had to be 3 minutes long! I went last and had the whole period to get even more nervous. When I finally got up there my knees were shaking and my speech ended up being one and a half minutes instead of three. Oh well, I'm just glad it's over with. My teacher says that we have to do another speech next semester but I don't want to think about it. this one was bad enough.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	18	(21)	4	(17)	0	0	
2	55	(56)	16	(70)	4	(100)	X
3	10	(12)	3	(13)	0	0	0
4	1	(01)	0	0	0	0	0
CONSENSUS SCORE:	2		2		2		2

Discussion of Sample C

Researcher/Trainer Response. Initially this paper seemed to deserve a score of 2, for being slightly superior to Sample B only because the tone or register and the sparseness of development seemed justified by the diary form. I was interested in whether this form would appeal to student raters and whether it would justify to experts the rater unelaborated recounting. The piece achieved verisimilitude--it sounded to me exactly like what a student might enter in a diary the night after such an experience, and it cleverly signalled that the experience was the first of many--a constraint the prompt suggests. There is not much more here in the way of vivid writing than in Sample B; the concrete images are mostly cliches--"scared out of my wits," "knees shaking." I was prepared to leave up to the rating groups the decision about whether to reward this paper with a better than bottom score because the language was appropriate for the diary form. Whatever decision was made would become binding for other such papers.

Student and Novice Teacher Rater Response. It was precisely the dilemma created by the fact of the

diary form that caused a wide division among students in evaluating this paper. Its defenders insisted that "it sounds just like a diary," and they went on to add that the description of how her knees shook and how her speech lasted only 1½ minutes really let us know how the author was feeling, as compared to the writer of Sample B. This sample was chosen as a representative 2 by all student scoring groups.

Novice teacher rater responses were distributed in much the same way as the students. They saw this paper as superior to Sample B for reasons similar to ones given by student raters.

Expert Teacher Rater Response. The experts did not make a distinction in quality between Samples B and C. This paper seemed a clear 2 to them, but so did the preceding one. Both of these seemed to them superior to Sample I, which they awarded a 1. They did not comment in discussion that the appropriateness of the language for a diary entry could be viewed as a special strength of the piece.

SAMPLE D

So I hear that your going to play a guitar for the first time. I know that at first you'll probably think that you're never going to be able to play the guitar successfully but don't let this discourage you. I felt the same way when I started but I kept on practicing and now it's easy for me. At first you may find that changing from chord to chord is the most difficult your fingers will tie up in knots running around themselves. The best way to overcome this problem is to use a very simple two or three finger chord and just strum the strings as you move each finger in unison up and down the six strings on the same fret and then do the same thing on different frets over the fret board; put your fingers in an A position on the second fret, strum a couple times, move your three fingers up to the fifth, fourth, and third strings, strum a couple more times, move back down to the A position, strum, and then move your whole hand up to the fifth fret and repeat. Soon you'll find playing the guitar just as comfortable as it is fun.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	4	(07)	2	(13)	0	0	0
2	30	(57)	5	(33)	1	(25)	(X)
3	22	(39)	8	(53)	3	(75)	X
4	1	(02)	0	0	0	0	0
CONSENSUS SCORE:	2/3		3		3		3

Discussion of Sample D

Researcher/Trainer Response. Samples D, E, F, G, K, L, M were set aside for second-round anchoring, as a check on where the cut-off should come between scores of 2 and 3. None of these seemed to be either bottom or top papers. Sample D was particularly interesting for training purposes, since it was clearly a response to Version B of the prompt which specified audience and rhetorical purpose more precisely. This student writer responded with a process paper, a set of directions on how to do something for the first time, as is quite appropriate given this version of the first-time prompt. The problem, of course, is that compared with accounts of personal experiences, this kind of paper seems dull to many readers--unless, of course, one is interested in learning to play the guitar. I was curious as to how students would evaluate this paper; would they verify my sense that these are good instructions--clear, accurate, helpful? I tentatively assigned this paper a 3, recognizing that it does not make exciting reading, but believing that it does well what it set out to do.

Student Rater Response. Students as a whole were not favorably impressed by this paper compared to others in the sample. The majority (52%) scored this a 2, although the feeling differed in the two classes that were given a second day of training. In each class students responded to my questions about the accuracy and helpfulness of these instructions qua instructions with fairly noncommittal comments such as, "They're all right. They're not wrong." Since few of these students attempted to write precise instructions for this test, they may have little understanding of the difficulty of the task; certainly, they had very few examples of such writing with which to compare this piece in order to discover its relative merit. This situation points up the problem of dealing with different kinds of writing in a holistic assessment and supports advocates of primary trait scoring who argue that different kinds of writing have different primary traits and cannot be fairly compared in a single ranking.

Expert Teacher Rater Response. Only one of these experienced raters considered giving this paper a 2, and she was uncertain. Given the anchors from the previous training session, the experts felt this paper was clearly as adequate as H and N (the mushroom paper and the airport paper) and better than B and C (first drunk and first speech in class), to which they had given 2's. Agreement was strong on a 3 ranking, so there was little discussion of the paper; no one complained that it was difficult to judge this informative, "how-to" paper against the more entertaining personal narratives.

Novice Teacher Rater Response. This group responded more positively to this piece, possibly recognizing the difficulty of writing clear instructions about a complex process. To some extent, judging from comments, these raters rewarded the writer for coherence, pleasant audience approach, personable voice, and complex but controlled syntax during the lengthy single-sentence instruction mid-paragraph. There was a strong consensus that this paper deserved a 3, but no one considered it a 4, while several thought it deserved less than 3.

SAMPLE E

Mom,

We are leaving here in about four hours. I had the most exciting vacation. This is the first time I have ever been away from the family alone. Even though my friend came with me, it's still different. You add so much security when we're in foreign places. I run into a distant stranger every second. I miss you and Dad because I never have to worry about a thing when you're near. I have learned alot this trip because I've had to worry about money, my car and eating. Don't misunderstand me, I need to do this. I feel as though I am just beginning to grow up. It's like a young sparrow leaving his mothers nest to go out into the world and defend his own life. It's a very scary feeling but a good one. I have enjoyed myself very much and it's been a great experience. Thanks for keeping me under your wing till I was ready to take flight. I look at this as one step towards independence. Thanks also for trusting me to be alone and being so understanding.

---your daughter---

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	2	(03)	0	0	0	0	0
2	14	(23)	4	(27)	3	(75)	0
3	29	( )	8	(53)	1	(25)	X
4	15	(25)	3	(20)	0	0	0
CONSENSUS SCORE:	3		3		2		3

Discussion of Sample E

Researcher/Trainer Response. In spite of the apparent cliché, "like a young sparrow leaving his mother's nest," and the general lack of specificity in the writing, I tentatively scored this paper a 3 rather than a 2 because I regarded it as doing effectively what it set out to do. Like Sample C (the diary entry) this writer creates an absolutely believable letter--one which might be written at the end of a journey, after other letters which were more like travelogues. This one would be intended to sum up the meaning of this extended "first time" experience. After hearing the experts discuss this paper, I became more aware of its weaknesses and see that, had we scored this paper 2 and lowered papers B and possibly C to scores of 1, we might have made a broader distribution of papers and been more precise in our use of 2 and 3 as scoring categories. Nevertheless, it is clear that this paper presents evaluation problems in comparison with other papers because, like Sample A, it deals with a so-called first

experience which was actually a series of experiences, and it deals with that experience on an abstract level. There is definitely more attempted here than in the "Dear Diary" of Sample C or in the rather flat, repetitious recounting of a first day at school in Sample G.

Student Rater Response. Students were divided fairly evenly over three scoring categories, with the bulk (48%) concentrating on a score of 3 for this paper, but with a strong minority (23%) giving it a 2 and an equally large group (25%) giving it a 4. Lower scoring students didn't have much to say about this paper: "I didn't like it. I don't believe she did that (traveled alone). There aren't enough details--she ought to tell something about a time when she wished her parents were there, but managed to get herself out anyway." The latter comment is, of course, a strong and valid criticism of this paper. Higher scoring students defended this paper as being a believable letter, answering the primary criticism of lack of details by saying, "That's not the kind of letter it is. She's writing at the end of the trip--she wrote that other stuff before." This view reflected my own view in choosing to give this paper a score of 3. Students had little difficulty in seeing that the paper was not strong enough for a top score, reaching consensus at 3. One interesting comment in favor of the paper was a compliment paid by one student in response to the metaphor of the sparrow leaving its mother's nest--reminding me again that what are cliches to adults may be fresh images for students.

Expert Teacher Rater Response. This group surprised me with 3 out of 4 readers giving this paper a 2 after having used the 3 category so freely. I asked them to defend their low scoring in this case. Line by line they pointed out the rather empty generalizations, some awkwardnesses. They were put off by the sparrow. They were not interested in considering the letter format or the special voice of the writer which might be assumed to result from the context of writing a letter to a parent. Given that context, they felt the student could or should have written more effectively--and that this paper was parallel in quality to Samples C and G rather than to Sample A, for example.

Novice Teacher Rater Response. The low raters in this group expressed the feeling that this entire paper was a cliché. They felt superior to this student's struggling effort to express emerging feelings of independence. Fortunately, a high rater pointed out what I had learned from the student responders: what seems trite to us may be very effective and new to students at this age. These readers agreed, however, that this writer is

not giving any reader new or deeper insights into what it means to gain independence; at best, we have a small portrait of how one adolescent responded with gratitude to being given responsibility. Certainly not a 4 paper, but the group was happy to agree on 3 instead of 2.

SAMPLE F

I can still smell the smoke. It lingers in our hair and clothing. The crackling climbing flames and dark smoke spread it seemed so quickly. Yet that came later. At first there were only those bright flickering lights and faint roaring noises, hardly threatening.

Mom woke us as Dad called the fire dep't. Their room was closest to the garage. She pounded on our doors shouting fire, telling us to get out of the house. Pepper reacted first. His barking and Mom's urgent tone at first confused me then her words sank in. Fire. Fire?

Anne & I scrambled out of our beds and grabbing Pepper hurried out of the house. The smoke made everything hazy, fantastic and yet real. We stayed at a friend's and talked to pass the time, catching up on the eight months in which we hadn't seen each other.

Dad, Mom & B stayed behind to save the furniture and personal belongings that they could. The fire consumed the garage, part of Mom & Dad's room and some of the kitchen & bathroom. When Ann

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	6	(10)	1	(07)	0	0	0
2	20	(34)	3	(20)	1	(25)	0
3	29	(49)	6	(40)	3	(75)	
4	4	(07)	5	(33)	0	0	0
CONSENSUS SCORE:	3		3		3		3

Discussion of Sample F

Researcher/Trainer Response. In all training for holistic assessment, I usually try to include one paper which is strongly written as far as it goes, but which appears unfinished. I want to alert the group to this type and arrive at a policy decision about how to score such papers. The alternatives are: (a) to imagine what the finished paper would be like, as if a page had simply been lost, scoring the paper on the available evidence of writing ability but not penalizing it for incompleteness; or, (b) treat it as a complete piece, with the absence of closure weighted into the holistic impression the piece makes. By either policy, I found the piece worthy of a 3 at least. The absence of closure prevented a score of 4 under policy (b), but the strength of the writing made it interesting enough reading to earn a good score. I could not award it a 4 under policy (a), judging its potential, because I felt it was unclear whether the student could end it well. The last paragraph to be written may well have petered out because the student

couldn't figure out where he was going or how to conclude what he had begun so dramatically.

Student Rater Response. This paper earned all four scores, the very lowest scorers basing their decision simply on the fact that the piece was unfinished, and that therefore the student hadn't really "done the assignment." It was easy to persuade these students to abandon this all-or-nothing formula for success and ask them to try to judge the writing that was there. A few students felt this paper included some of the best writing of all the anchors (especially the first paragraph), and so they wanted to award it a 4. They agreed to a lower score when reminded that after all the writer had not finished and we don't really know if he could have finished well. Finally, the debate centered on whether the paper deserved a 2 or a 3. Because the aim of the training session is to achieve agreement, not to establish absolutely the merits of each paper, discussion was not prolonged. I presented my view that there was sufficient writing of good quality, that the paper was rewarding to read even without an ending, and that it was stronger and more interesting than the other papers we had scored as 2's. Students seemed comfortable with this resolution and were generally willing to take the view that unfinished papers can be judged for what they offer, without trying to speculate about how well they might have been completed had the writer been given more time.

Expert Teacher Rater Response. The experts agreed in general with the novice teacher rater and my response, discussing the extent to which the absence of an ending interfered with the reader's ability to enjoy this piece and understand the significance of that "first time" experience. One expert felt the lack of ending to be greatly detrimental, resulting from a lack of overall planning and a sense of purpose in writing the essay. All readers felt this student had excellent narrative skills and a real sense of how to begin a dramatic account.

Novice Teacher Rater Response. There was a wide range of responses, in spite of previous training, because of the missing ending. This group, however, produced a larger percentage of top scores than the student raters had because they were more impressed by the rather effective writing of the opening paragraphs. This group produced good criticisms of the piece, however, saying that the sudden loss of focus and the telescoping of events at the end of the third paragraph suggest the writer had abandoned the moment worth writing about and was going to have difficulty finding good closure--even with adequate time to write.

SAMPLE G

I remember what my first day of high school was like perfectly. I remember how frightened I was that I would do something stupid. I felt like I was the only person in the school who felt lost and that everyone else knew exactly what they were doing. Instead of going to my locker between each class, I brought all of my books with me. I guess I was afraid that I wouldn't have enough time to go all the way to my locker and then try and find my next period class. Before each class I would practically pray that there would be someone inside that I knew. I remember thinking that I would never make any new friends. I felt so nervous the whole. I'm so happy that day is gone forever. I will never forget how nervous I was and what a relief it was when I heard the 6th period bell ring to let us go home.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	2	(03)	3	(20)	0	0	0
2	7 + 24	(52)	12	(80)	4	(100)	X
3	22 + 4	(43)	0	0	0	0	0
4	1	(02)	0	0	0	0	0
CONSENSUS SCORE:	2/3		2		2		2

Discussion of Sample G

Researcher/Trainer Response. This sample is much like Samples B and C in level of detail and scope--perhaps a trifle stronger than these, but certainly not up to Sample H (mushrooms) in overall impact. The piece is unredeemed by the kind of lively pointed ending that saves Sample N (airport). It seemed a good paper to use to stabilize scorers' concepts of a 2 paper, having a little more substance and coherence than the papers representing the 1 score (Samples I and possibly B).

Student Rater Response. Students were split between the scores 2 and 3. The appeal of this paper was strong and apparently based in part on its subject: defenders spoke of how vividly and convincingly the writer had captured an experience they recognized. "It's just like that," was a common comment. These defenders rightly saw more details than in weaker papers and were not easily convinced that the paper was not strong enough for a 3 until they were asked to re-read Samples A and H. Even then, some maintained that this paper was just as good, and I allowed the difference in opinion to stand, pointing out that every assessment yields some borderline papers, and that the only scorers

who were really off-base were the 1's and 4's. (This kind of training is dangerous, of course, since it encourages scorers to stick to middle scores and not trust themselves on the extremes.)

Expert Teacher Rater Response. These four raters were unanimous in their recognition of a lower half paper with some redeeming strengths and agreed with the trainer's assessment. There was no discussion of this paper.

Novice Teacher Rater Response. By the second training session, this group was better at spotting lower half papers and remaining clear in the reasons for their choice.

SAMPLE H

Mushrooms

Two years ago my Spanish teacher took us out to lunch. We went to Thunder Mug. It's a favorite Pizza Parlor my old junior high school used to go to. Before we left we had to decide what kind of Pizza we wanted to get, so my teacher could call in and order it to be ready when we got there. There were five people in my group. We voted on what kind of pizza we would order. All four of them except for me, wanted a mushroom pizza. I didn't want to argue about it so, I went along with it. But I wouldn't eat the mushrooms. I'd take them off. When we got there the Pizza was ready. I saw all throughs goobs of mushrooms on it, and thought I would get sick. We found a table and they started eating. I took a piece. I looked down at it and all thoughts mushrooms. I thought of the last time I had had mushrooms. (those sick gray things that tasted like rubber (all mushy) I started to pick them off. Every body sat there staring at me, and I felt like a fool, so I put the mushrooms back on, closed my eyes and took a bite (try not to taste anything.). I tasted good. I couldn't believe it. I took another bite, and it tasted even better. Soon I had downed more then my share of the pizza, so I stopped eating. From that day on, all my Pizza's have tons of mushrooms on them. I now put mushrooms (in) on just about everything. I eat them on Pizzas, on steak, sauteed. (next to just about any meat, in salad, mixed in vegetables. I don't see how I couldn't have known how good they are. I'm glad I tried to eat them two year ago, or I still would not know how good they are.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	3	(05)	0	0	0	0	0
2	13	(22)	6	(40)	1	(25)	(X)
3	32	(54)	7	(47)	2	(50)	X
4	11	(19)	2	(13)	1	(25)	0
CONSENSUS SCORE:	3		3		3		3

Discussion of Sample H

Researcher/Trainer Response. In the initial selection of anchors, I considered this piece inferior to Sample A, but superior to Samples B, C and I--and about equivalent to D and L. I anticipated eating B, C, I, as 1 papers, and rating Sample H (with D and L) as 2 papers, saving the score of 3 for papers as good as Sample A. I found the straight narrative style of this paper rather flat; I was bored at the beginning and reluctant to keep on reading. Only during the training of rating groups did I begin to value this paper for the clear impact of its simple anecdote, and I found myself persuaded to raise the score of this paper to 3. Unfortunately, when that shift was accomplished, a good many other papers which were weaker than

Sample A, our first well-established 3 paper, were also scored 3. This paper, then, and its popularity, contributed to expanding the 3 category at the expense of other scoring categories.

Student Rater Response. There was no arguing with the appeal of this paper--almost as many students wanted to give it a top score as wanted to give it a 2, with the bulk of scorers preferring a score of 3. They couldn't argue about the comparative syntactic resources of Samples A and H--all they knew was that they had enjoyed these papers about equally, and that both were "more fun" than most of the other papers in the sample. I acknowledged their certainty; their taste, after all, was an important measure of the success of this piece of writing. The few students who gave this paper the poorest score admitted that they hadn't finished reading it--apparently the opening had affected them as it had me.

Expert Teacher Rater Response. These four readers initially were quite divided, eventually choosing a 3 by consensus, as did the other groups.

Novice Teacher Rater Response. This group, like the student raters, simply enjoyed this account, reading into it a nice lesson on how we learn to like something when we are forced to try it. One pointed out that the straightforward narrative style was a successful way of organizing a response to the prompt; the reader always knew what was going on and didn't have to make any leaps. The cumulative effect of the piece was strong, whatever one thought about particular sentences or phrases. The central part of the essay, with its freely and rather carelessly used parentheses, struck many readers as delightful; they liked this student's "voice."

SAMPLE I

When I was about five years old I put some insect poison in my mouth not knowing that it was poison. I later found out that it was poison, and became very scared. I remember hiding by myself wondering if I was going to die. After a while I was still feeling allright so I came inside, went about my normal routine, and soon after forgot the incident. It embarassed me so much that I never told anyone about it, and I think this is why I still remember it so well

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	31	(59)	10	(67)	4	(100)	X
2	18	(35)	4	(27)	0	0	
3	3	(06)	1	(07)	0	0	0
4	0	0	0	0	0	0	0
CONSENSUS SCORE:	1		1		1		1

Discussion of Sample I

Researcher/Trainer Response. This paper had an appeal--a kind of simple dignity. This writer, like the one in Sample B, is facing the problem of writing interestingly about not experiencing symptoms he could report in vivid detail. He had to report what had happened in his imagination--how it felt to think he was going to die. Hemingway very neatly avoids this problem in "A Day's Wait" by telling the story through a father's eyes; the child's behavior is described as simply as in this writer's account.

Nonetheless, this writer clearly does not attempt much and is satisfied with a bare recounting of the incident. As far as it goes, it would be a perfectly acceptable oral response to an oral request, but it does not recognize the contract implied in a written test, a contract that might be described as "show off your writing" (see Sample J), or less perjoratively as, "write so that we can relive the experience with you--write as if you are a writer, not a test taker or a respondent to a simple question in an oral dyad."

Student and Novice Teacher Rater Response.

Student rankings were somewhat more spread on this paper than on Sample B. Again, the recognition of an experience which has been told simply and clearly and which itself carries an impact appealed to many students. The preceding discussion of Sample B, which called for more details, inspired many students to defend their low scores for this paper (as did the experts) by pointing out simply the lack of elaboration, the need for "showing, not telling." Novice teacher raters had similar reactions.

Expert Teacher Rater Response. The experts were unanimous in citing this paper as the anchor for the lowest possible score. As with Samples B and C, the unanimity invited little discussion. This paper was a recognizable type to the readers from their experiences in reading for ETS: the writer simply doesn't attempt enough.

SAMPLE J

Dear Rachel,

I am so excited that you are coming out to visit me this summer. The first thing we are going to do is go to Marriot's Great America and ride the Tidal Wave. Remembering that you've never been on a roller coaster before, I promise I'll be patient with all your screams and pleas for deferral. But don't worry because I reacted the same way the first time I soared through the skies upside down. In fact, my first experience on that death-defying machine will always swirl through my mind as the roller coaster whooshed through space. Accompanied by three other friends, I had bravely approached the long line extending from the mock ship. Somehow, waiting the forty-five minutes it took to reach the front of the queue was far easier than the last 60 seconds before my turn arrived. Up close, the noise vibrated the wooden platform more so than the actual rush of acceleration. My face changed a different hue every breath I took, which occurred rather frequently, as I stepped unwillingly into the metal death chamber. Hastily the amusement park employees recited their instructions unheard by me as I recited my prayers. Before my "Please God" had been completed a giant palm swatted the back of the train and hurled it forward along the ominously constructed tracks. My head yanked backwards against the head rest; my eyes clamped shut; and my sweaty palms acquired a grip impossible under normal circumstances. Within a matter of seconds I could feel an unaccountable force pumping all my blood into my head. Opening my eyes for an explanation, I discovered a transformation in the earth's terrain--I was upside down. Pushed through a muggy summer day, the sweat dripping down my face flew off at the corners. And the next thing I knew, the mechanical caterpillar jerked to a halt, and I was hustled dizzy and dazed from the train. Smiling, I walked away saying, "Boy that was fun let's do it again."

You'll love it Rachel. Can't wait to see you.

Love  
J----

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	0	0	0	0	0	0	0
2	2	(02)	0	0	0	0	0
3	27	(33)	7	(30)	0	0	0
4	54	(65)	16	(70)	4	(100)	0
CONSENSUS SCORE:	4	--	4		4		4

Discussion of Sample J

Researcher/Trainer Response. I chose this paper as a nearly ideal response to the prompt, representing without question the top score possible. I was only concerned whether there would be enough papers in the sample that so successfully responded to the task. If this paper were too far superior to most others, it would create too narrow a scoring category for the

score of 4, eliminating many other papers which should be considered top papers if the reading were to obtain a good spread of papers across the scoring range.

J has addressed her narrative to "someone who is about to have a similar experience," providing a realistic rhetorical context, choosing an appropriate form (a letter), and working within these constraints to write a particularly vivid account of her own first experience on a roller coaster. The flatness of the first few sentences, and the overwriting of sentence 6, I attributed to problems in creating the setting--problems prompted by the task. As creator of the task, I was quite willing to read past the writing problems that seemed to result from the constraints I put on the writer (ostensibly to help her write better!), privately congratulating her on making such a good attempt to couch her anecdote in the form invited by the prompt. I skimmed these first sentences to get to the "real" purpose of the assignment, her recounting of her own first-time experience. I was soon impressed by, "Somehow, waiting the forty-five minutes it took me to reach the front of the queue was far easier than the last 60 seconds..." and, "the employees recited their instructions unheard by me as I recited my prayers..." I read on, looking for a good solid physical sensation, and I found it, as the "giant palm swatted the back of the train..." This student had fulfilled my expectations, and continued beautifully to do so for the rest of the essay cum letter.

Student Rater Response. I was startled to find that students did not universally recognize this paper as the apex of achievement, and they had good reasons for their reservations. The most striking comment, offered as a defense of the low mark given by a responsible student in a college prep literature class, was simply: "It sounds too much like an English theme." Other students spelled out this criticism, citing line and phrase (especially in the beginning of the letter), complaining that no one would ever write a letter like that. I believe that they were reacting negatively to two different problems in the writing of J on this occasion. The primary problem is the one created by the prompt, "write for someone who is about to have a similar experience." The students were more harsh in their judgment. Essentially, they said that if this writer wanted them to believe this was a letter to a friend, she couldn't write it like a composition to an English teacher. Fortunately, for purposes of training raters to reach agreement, there was at least one student in each class who argued that it could be a letter to a friend. The classroom teacher intervened to inquire,

"Wouldn't you like to get letters this vivid from your friends?" and received a generally positive response. In other words, we can probably train students to recognize that English teachers are going to value and give high scores to good English-teacher-approved writing even in the context of an artificial letter format and even if the students think the writing is unbelievable in that context.

A second criticism the students seemed to be making is one they had no word for: "overwriting" is how the novice teacher raters put it. "My first experience on that death-defying machine will always swirl through my mind as the roller coaster whooshed through space" is a sentence that gets more points for effort than for effect.

Expert Teacher Rater Response. The four expert raters unanimously chose this paper as representing the top score, wasted little time discussing it, and did in fact fall into the trap the researcher/trainer was concerned about: finding few other papers in the sample that so clearly and adequately responded to the test-maker's intentions and to these trained raters' criteria for success. The level of diction and syntax, the flawless mechanics, the images, the telling details, the play on "recited...recited," the cohering devices (moving from "my prayers..." to "Before my 'Please God' had been completed..."), the device of setting the anecdote in a letter format as a solution to the audience constraint--and then making sure that at least the introduction and conclusion sounded like a real letter--all of these qualities in one piece of writing make this an exceptional paper.

Novice Teacher Rater Response. All the novice teachers recognized that this was an upper half paper, but they split very much like the student raters over whether it was really a top paper, with 30% complaining that it wasn't a "believable letter to a friend" or that it was overwritten. One reader put it, "I just don't like her very much," and another explained, "She's showing off her writing ability, rather than really writing about the experience." An interesting commentary on what a writing test performance evokes.

SAMPLE K

It's funny how some people think more about the things they dislike than those that please them; one'd think every body would prefer dwelling on their virtues and advantages than their short comings. I dislike writing under given topics. I live for the present. Thusly from out of a digression I conform to the subject and describe the first time I wrote for an experiment.

The day had been mundane, the weather warm, my lunch to large and none too good; upon entering my English class a stapled set of papers was thrust on me and I was sent into an adjacent room to do whatever it was the sheets dictated.

Now, to make on thing clear: My ego is at odds with my common sense. The latter would dictate my avoiding the wrath of Miss McNelly by buckling down and doing my best; the former compells me to sneer and retreat... unjustifiably. This paper was begun only after twenty minutes of battle had passed between the two factions of my consciousness, for in order to write up to my standards, I would have to say something important to me... which I would not reveal simply for the sake of an experiment. A fault of mine, I suppose...I am too reticent towards baring my innermost soul to those people who simply have every business knowing; my peers and my teachers. So, I take a sneaky-panther cop-out--- and here you have your data.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	2	(03)	1	(07)	0	0	0
2	18	(29)	7	(30)	0	0	0
3	21	(34)	6	(26)	3	(75)	X
4	21	(34)	9	(39)	1	(25)	0
CONSENSUS SCORE:	3/4		3		3		3

Discussion of Sample K

Researcher/Trainer Response. Perhaps one of the most intriguing papers in the sample, this one was included as an anchor because I wanted student raters to see how a writer might argue with the test situation, protesting his "first-time" role as a guinea pig while still technically "conforming to the subject" and writing an acceptable composition. The writer's agonies, however, do take their toll rhetorically: I would rate the paper a 3.

Student Rater Response. Most students recognized the strength of this paper--a good number, however, gave it low scores because they felt it didn't make sense or wasn't an adequate response to the assignment. Fortunately, in each class the criticisms were heard first, inspiring an articulate and spirited defense by some

student who sympathized with the writer's dilemma and was enormously impressed with his method of solving it. These defenders were often unwilling to accept my verdict that the paper had flaws that would make it a 3, and so I let it stand for this group as a borderline, 3/4 paper.

Expert Teacher Rater Response. All four experts appreciated fully what this student had achieved, but three agreed with my first assessment that the roughness, especially of the opening paragraph, and the connections the reader had to provide throughout, made this less than a top score.

Novice Teacher Rater Response. The novice teachers were as enthusiastic as the positively impressed student readers, with critical comments coming primarily from those who seemed not to have read the paper carefully enough to recognize what had been attempted and achieved.

SAMPLE L

In my youth I hardly ever got sick, but when I did, I grew very sick. Such was the case when I first experienced a hospital. My temperature had climbed far above normal and wouldn't regress. I remember most vividly the trip through the halls where people from all walks of life could be seen. Illness, the only thing that they had in common, brought them together, each of them unconcerned with any other. I had a hard time not feeling sorry for them until I was the one manipulated by the nurses. Everyone rushed, while I hardly could move. Hospitals, which cure one's body, also take a bit of one's sanity.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(one reader)
1	0	0	0	0	0	0	0
2	3	(11)	12	(52)	2	(50)	X
3	18	(67)	10	(43)	2	(50)	0
4	6	(22)	1	(04)	0		0
CONSENSUS SCORE:	3		2		3		2

Discussion of Sample L

Researcher/Trainer Response. I chose this paper as an example of a good 2 score, remaining at too abstract a level, lacking development, but with nice syntax and a last line which can be read as cohering all that goes before.

Student Rater Response. Students were more impressed than I was by this paper, and judged it a 3 in the one class which had time to read and rate it as a group. They argued strongly with me, and I allowed them this one as being a little more imaginative and effective than such papers as Samples B, C, or G. I wanted to encourage them not to consider length as a sine qua non for high quality, and it seemed useful in training to allow one short paper a score of 3 based on other merits.

Expert Teacher Rater Response. These four raters were divided evenly on this paper; two of them reflected the response of myself and the novice teachers, two others responded as the students had. Defenders read the paper aloud pointing out that it was tight, consistent, and said everything that needed to be said to support that last line, which they found quite effective.

327

Detractors, myself included, complained that the long sentence quoted by the novice teachers was difficult to understand, rendering unclear the point of view of the whole piece. Defenders in turn accused low scorers of careless reading. The low scorers replied with maxims about the responsibility of the writer to help the reader see. Whenever a paper divides a group along these lines, with some readers seeing it as full of meaning and others believing the meaning has to be read in, it is usually best to abandon the sample as an anchor, thus, "agreeing to disagree." In this case, however, the expert group reached consensus, choosing to regard this as a 3 rather than as borderline. The consensus did reflect the tendency of this group to score a paper 3 when in doubt.

Novice Teacher Rater Response. This group was somewhat inclined, like myself, to score this paper down, not only for brevity, but for unreadability. They responded negatively, as I had, to the sentence: "I had a hard time not feeling sorry for them until I was the one manipulated by the nurses."

2.31

SAMPLE M

Art has different meanings to different people. Art for me isn't what it is for another; the meaning can change even for each individual work simply by being interpreted differently in the mind of the experienter.

M.C. Escher was one of the most brilliant and talented artist of our time. He died in 1972, but his works live on with the image of his mind forever recorded on them. To look at one of his works (most being ink paintings) is to look into a strange dimension somehow hidden from the rest of the world...they're alive with ideas, every one of them. It was the dynamic aspect of them which I like and have liked ever since I first saw one about five years ago. Many find his painting strange and illogical, not pleasing to the eye. I find them just the opposite. They contain increadable beauty and comfort, both of which I saw and felt from the start. In a world of hazard and discomfort, Escher is a refreshing escape into another world, one of those escapes we all need from time to time.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(One reader)
1	(X)		(X)		0	0	
2	(X)		(X)		0	0	
3	(X)		(X)		2	(50)	(X)
4	X		(X)		2	(50)	
CONSENSUS SCORE:	EVASION (4)		EVASION (3)		3/4		EVASION (3)

X = first choice score, no count made  
(X) = strong second choice score, no count made

Discussion of Sample M

Researcher/Trainer Response. A problematic paper, this essay represents a common dilemma for holistic scorers. At first glance, I read it as an evasion of sorts. If I had encountered it on a College Board English Composition test, I might have thought it a canned essay, a favorite collection of abstractions used successfully on another occasion and tailored slightly to appear to fit the topic. However, I altered my position and came to recognize that it is equally likely that this paper is a fairly sophisticated and ambitious attempt to deal with the topic, a sincere response to the Version B admonition to write for someone about to have a similar experience. Hence, the writer's introductory sentences point out that different people have different responses to art. Like a few others who responded to Version B, this student interpreted the instructions, quite validly, as requiring writers to address themselves to preparing the reader for an initial experience, rather than simply recounting the writer's own initial experience.

scored but should be turned over to the trainer or the teacher. This is done to avoid giving such papers misleadingly low scores.

Finally, one student pointed out that even if we thought the paper was on the subject, it should get a score of 3 not 4 because "it would be a better paper if the person really had told us something about his first experience--so we could see the paintings." Ultimately, this was the view that I took of the paper.

Expert Teacher Rater Response. The experts were adamant that this paper was not an evasion, and scored it with little disagreement as a top half paper. The group split 50/50, with two readers giving it a 4 and two giving it a 3. Readers were not able to reconcile this difference; staunch defenders of the maturity involved in what the student attempted and the quality of his abstractions felt that this paper was far superior to papers such as A and H, and as good as J, although different in kind. The other two readers agreed with the researcher/trainer that the absence of supporting details weakened the paper, although they recognized the worth of what the student was attempting and the superior quality of the abstractions, including the introduction and conclusion.

Novice Teacher Rater Response. Novice teachers responded to this paper very much as the student raters had: they admired it as a good piece of writing, on the whole, with a few raters giving it low scores for failing to deal with the topic. Again I introduced the option of labeling a paper as an "evasion" to avoid using a low score for a well written paper. As with student raters, the use of this sample paper as an example of an evasion did not cause the teacher novices to score any other paper this way. Strictly speaking, then, if this paper is not considered an evasion, then there were no evasions in the entire sample of student writing.

With this in mind, a main shortcoming of the paper is the lack of concrete details to help us visualize what it is Escher does--in this case an enormously difficult task, possibly beyond the student's powers. But equally possible is the fact that in thirty minutes, the student found insufficient time to do more than work out the fairly strong generalities that express his feelings about Escher. If we use as our rubric the standard directions to teacher scorers at Sir Francis Drake High School in Marin County, California: "Treat each paper with respect. The student should be rewarded for what he does well," we might find ourselves agreeing with the one expert scorer who awarded this paper a 4. I decided that my initial impression, that this paper was an evasion, was in error, and I believe that with a more careful reading I would have scored this paper 3. The claims, albeit fascinating ("they're alive with ideas, every one of them"), need some attempt at substantiation--description either of the paintings or of one or two of the ideas they are "alive with." However difficult this might be to do, the student set himself this task when he tackled this subject, and from the rhetorical perspective, this piece remains unfinished. The implication for holistically scored time writing tests is, of course, that part of the student's task is to select material that can be turned into a fairly finished piece in very little time. Some subjects are too big and too important to be dealt with on such occasions. And yet, it will be the most able students who fall into this trap, a fact which mitigates against short timed tests.

Student Rater Response. I took an informal vote of initial responses to this prompt but I did not record them, unfortunately, since I expected to point out to students that this paper constituted an evasion. In all classes, however, I observed an inverted curve in the response; that is, students who saw the paper as fulfilling the topic gave it one of the top scores (over half the students in each class gave this essay a 4), with a few students giving it the lowest score as a way of saying that it did not satisfy the assignment. Comments in favor of the top score for this paper showed that students generally were impressed and attracted by the subject chosen, by the abstract concepts, and by the mature voice. When comments were requested from the students who scored this essay low, the response was simply, "It doesn't even deal with the topic!" No students were able to show, as the experts later did, how the paper could actually be read as very much on topic; all students appeared to accept the trainer's initial instruction to regard this paper as a model of evasion, understanding that such papers should not be

SAMPLE N

The first time I walked through an airport and watched the planes lift off while staring through the immense plate glass windows was really exciting. The hustle and bustle of the people, the noise of the jets on the slick floors of the terminal were just driving me crazy. There were so many different types of people, with slanted eyes. Some people couldn't talk correctly, they spoke in a jibberish sort of way. I saw more types of people than I could imagine while I was sliding back and forth across the slick floors. Mother thought they were dirty, but she was wrong they weren't dirty but there sure were a lot of little rolled up pieces of paper with lipstick on them. The best was yet to come. With my nose pressed tightly to the window I could see the taxiing around the runways. I could even see a few taking off, first the front and then the back, floating through the air as easily as I could imagine. The excitement of the airport is gone now. I've seen the planes take off a hundred times, and I've had my fill of slick floors and cigarette butts, and the people who were different I now see every day. The planes are still a little special though because they are the ticket to new firsts.

	STUDENTS		TEACHER NOVICES		EXPERTS		RESEARCHER
	n	%	n	%	n	%	(One reader)
1	0	0	0	0	0	0	0
2	3	(11)	12	(52)	2	(50)	0
3	18	(18)	10	(43)	2	(50)	X
4	6	(22)	1	(04)	0	0	0
CONSENSUS SCORE:	3		2/3		3-		3

Discussion of Sample N

A Note from the Researcher/Trainer. Sample N was read by only 27 students, having been included in the first training packet as the last paper of seven. In all but one classroom, discussion of the other six anchors took so much time that scoring was begun without tabulating responses to this paper. Considering the spread of responses from other scoring groups, I regret not having recorded more student responses. I found this paper parallel in general quality to Sample H, the "mushroom" paper, and wavered--as did the expert raters--on whether to award it a 2 or a 3. This is not because I don't see it as clearly superior to certain other papers and clearly inferior to still others, but because it is not clear to me where the cutoff should come between 2 and 3 for this sample of papers. Like the expert raters, I have read the California High School Proficiency Exam for ETS and have internalized a set of expectations that leads me to see these samples (N and H) as relatively good writing--strong 3's

on a 5-point scale. But the scale in this study was only four points, and there were virtually no papers in this sample equivalent to the papers that earn scores of 1 on the CHSPE. This would tend to lower the level of papers such as this one to a score of 2.

Apart from the debate about where to draw the line between 2 and 3, the paper itself creates some debate. I read the paper as adopting the voice and perspective of a child during the bulk of the first paragraph, then returning to an older perspective (and hence voice) in the final paragraph, with strikingly effective closure. The argument between readers who placed this paper in the upper half and those who found it typical of lower half papers centered on whether the childish language of the middle passage ("so many different types of people, with slanted eyes....Mother thought they were dirty but she was wrong they weren't dirty...") was indicative of an undeveloped writing style or was an attempt to assume a tone appropriate to the child who was sliding around on the slick floors picking up cigarette butts. In the last analysis, all scorers agreed that the details in this paper and the fine ending deserved a higher score than papers they had agreed to call 2, such as C. It might be useful to consider this paper typical of a 2/3 split judgment.

Chapter 3

BEYOND THE FIELD TEST OF WRITING PROMPTS:  
A PRELIMINARY INVESTIGATION OF THE  
NATURE OF RESPONSE VARIATION TO  
SELECTED FIELD-TESTED PROMPTS

Gerald Camp, Charles Kinzer, and Sandra Murphy

Part I: The Effects of Assessment Prompt and  
Response Variables on Holistic Score:  
A Pilot Study and Validation of an  
Analysis Technique

Charles Kinzer and Sandra Murphy

Part II: A Field Test of Eight Prompts: Santa  
Clara County Writing Assessment\*

Gerald Camp

\*A Bay Area Writing Project Study conducted on  
behalf of the Santa Clara County AB 65 Consortium,  
which provided the data base for the study in  
Part I.

THE EFFECTS OF ASSESSMENT PROMPT AND RESPONSE  
VARIABLES ON HOLISTIC SCORE: A PILOT STUDY  
AND VALIDATION OF AN ANALYSIS TECHNIQUE

Charles Kinzer and Sandra Murphy

1.0 Introduction

Holistic scoring procedures for writing have been widely accepted and adopted by both researchers and practitioners in the field. Traditionally, holistic scoring has been used to rank papers for placement and competency testing purposes. Following such rankings, some researchers (Cooper, Cherry, Gerber, Fleischer, Copley, and Sartisky, 1980) have analyzed characteristic features of the written product in the holistically scored sample. However, research exploring features of the assessment writing prompt which may influence holistic scores is a relatively new phenomenon.

Freedman (1977), in an examination of the effects of reader, essay and environmental variables on holistic scores, found that essay variables contributed most to the variance in the scores. Unexpectedly, topic effects were also found. The focus of the study reported here is upon possible relationships between assessment prompts, written products, and holistic scores.

As part of a larger NIE project, this study is one of a series that examines writing task variables in holistic writing assessments. This is a preliminary effort to determine ways of identifying effects of prompt and response variables in relation to holistic scores.

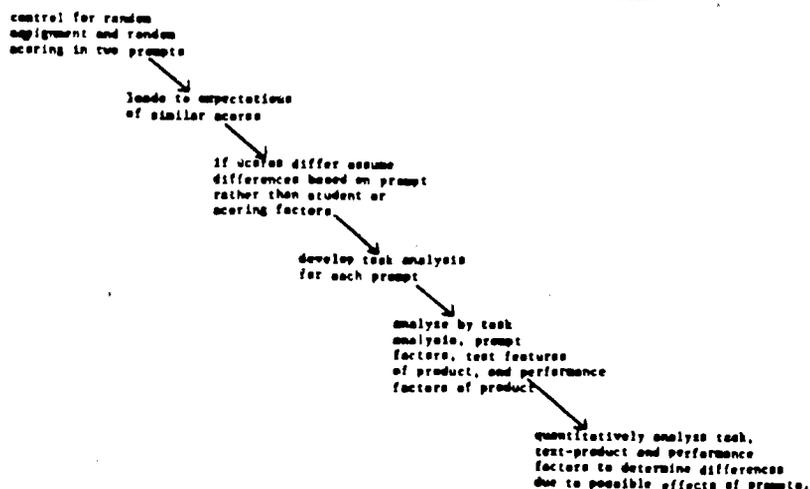
This study therefore had two major goals:

1. To investigate prompt effects and response variables which might result in discrepancies in holistic scores.
2. To validate an analysis system for looking within specific prompts.

A fairly rigorous design which would incorporate prompts, scores and written products while controlling for extraneous variables is shown in Figure 1. Figure 1 also notes the basic assumptions upon which this study is based.

FIGURE 1

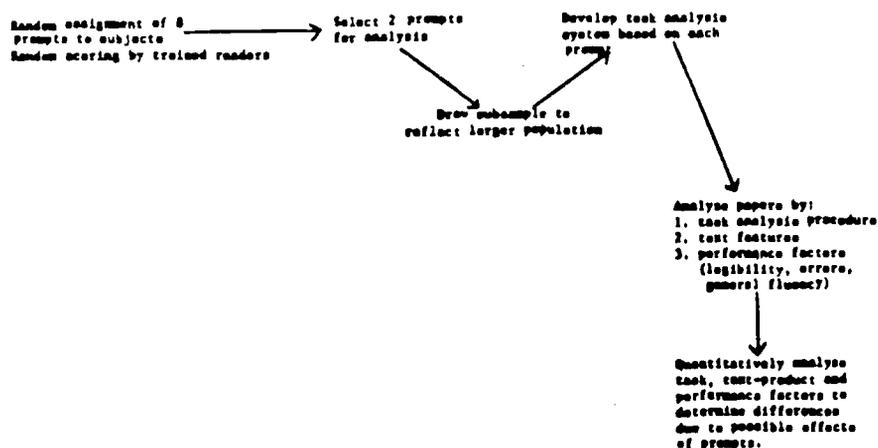
General Framework and Premise for the Study



The specific design for the study parallels that shown in Figure 1 (see Figure 2). The modifications arise from this study's evolving into a larger, Bay Area Writing Project assessment.

FIGURE 2

Experimental Design



## 2.0 The Data Base

In 1979, the Bay Area Writing Project and the Santa Clara County AB65 Secondary Consortium developed a set of writing topics, or prompts, which were to become part of the Consortium's bank of evaluation instruments. The prompts were developed by a subcommittee made up of writing teachers from Santa Clara County elementary and secondary schools. Prompts from several sources were examined and members of the committee also contributed

topics they had used in their classrooms or districts. Twelve prompts were initially drafted and revised. (See Part II of this report for a full description of the data base.)

The final prompts developed by the committee were field tested in several districts throughout the country. Eight prompts were tested.

1. Imagine that you are trapped in one of the following situations and you cannot escape for several hours. Tell what you do to escape or to pass the time until you are rescued.
  - a. an elevator
  - b. an abandoned mine
  - c. a locked room
  - d. a life boat
2. If you had to choose to be something other than a human being, what would you choose? Give your reader some idea of what it would be like to be that form, and why you chose it.
3. Think of a personal experience that has in some way changed your life: a particular event; a person; a place you have visited; a book you have read; a film, television show, or play you have seen. Describe the experience and explain why it was an important one for you.
4. While visiting in your grandmother's house you find a large trunk. Propped on a chair beside the trunk is a card, and on it in your grandmother's handwriting are the words, "To my only grandchild: Inside this trunk you will find what you have always wanted." Describe the gift and how you will use it.
5. Think of a friend, real or imaginary, that you had when you were younger. Describe something you and this friend did together; try to show your reader the kind of person this friend was and why you chose this person as a friend.
6. The school newspaper has asked students to submit suggestions about how to improve the school. Think of ONE and only one problem which you would like to see solved to make life at school better for you and others. In a letter to the school newspaper describe the problem and tell how you would like to see it solved.
7. Employment agencies can be very helpful in finding you a job if they know what job you want and are qualified for. Write a letter

to the ABC Employment Agency, 204 East Main Street, Bay City, CA 94606. Explain what job you would like and what your qualifications are. Explain the reasons you think you should be hired and give examples of your strengths. Include references.

8. You have been sent a bill charging you for two record albums you did not purchase. Write a letter explaining the problem so that someone in the billing department can correct the error.

The bill is from Melodie Record Club,  
444 Tune Street, Nashville, Tenn. 78056

The records charged to your account are:  
Kiss Presents Opera Highlights and  
Led Zepplin Country Hits

Your name and address are:  
Chris Brown  
456 Main Street  
Bay City, CA 92001  
Your account number is:  
35-800-21

Data were obtained on the performance of each prompt in typical county schools identified as low, mid-range, and high. The school categories were based on student performance on other measures of their writing ability. Schools were also categorized by percentile ranking on CTBS, and the county-wide proportion of schools which fell into each category was determined.

The mid-range sample formed the data base for the study reported here. In this sample eight classes, divided among three schools, were tested. One school fell into the 34-44 percentile range, one into the 45-55 percentile range, and one into the 56-66 percentile range. These schools were further categorized as low, medium, and high respectively. The eight prompts were tested in one "low," two "average" and one "high" ability class.

In the three mid-range schools, between 23 and 30 papers were obtained from each class. The eight prompts were randomly assigned to the students in each class. Testing was done during a five-day period. Students were given 30 minutes to write in response to the given prompt. The data from the above field test of prompts were made available to the investigators for the study reported here. As noted, identification of the effects of the prompts, and the development of systems for investigating and analyzing such effects are the major

goals of this study. In keeping with the prospective design noted in Figure 1, two prompts were considered ideal for the purposes of this study. Mean scores between prompts 5 and 6 were significantly different. Such a difference was unexpected due to the randomization in assignment of prompts and controls in the scoring procedures. Thus, the papers resulting from responses to these two prompts were selected for analysis in this study. In short, prompts 5 and 6 were selected because they showed the widest variation in average holistic score. Means and variances for the two topics are shown in Table 1.\*

The sub-sample selected for analysis was drawn from the entire pool of responses to prompts 5 and 6. To ensure that the sub-sample was representative of the total data base, the data base was divided into number of papers per score. Within each score, papers were randomly selected so that overall percentages of papers per score were equivalent across the total response set and the sample to be analyzed. For example, if 20% of the total data set received a score of 8, then the sample was selected so that 20% of the total to be analyzed received a score of 8. In this way, the sample was both randomly selected and representative of the data set, allowing generalization to the population from which the original data was gathered. Forty papers, twenty each from the set of responses to prompts 5 and 6, were selected for analysis (see Table 2).

### 3.0 The Analysis System

After the sample was selected, a task analysis system was developed to provide a means for identifying the salient aspects of the prompts and student responses. The prompt-related task analysis system was based specifically on the two prompts. The analysis system should be viewed as an exploratory methodology. It was constructed to allow examination of performance and prompt factors in relation to holistic score both within and across prompts. Global analysis variables in the response analysis included legibility, general fluency in terms of the amount of writing produced, use of cohesive devices in the development of ideas, and errors related to grammatical and written language conventions.

#### 3.1 Response-based Variables

##### 3.1.1 Legibility. A measure of legibility was

---

\*All tables are at the end of this report. <sup>24</sup>

obtained through an analysis procedure previously employed by Keech and Blickhahn (1981). Individual papers were ranked on a scale of 1 to 4 by two of the investigators. Factors used for the legibility rankings are shown in Table 3. Each legibility score was based upon two independent rankings.

3.1.2 Cohesion. A measure of idea development was obtained through the use of an adaptation of Halliday and Hasan's technique for analyzing cohesion in English (Halliday and Hasan, 1976). When groups of words within a text were related to the same idea, they were treated as textual "chains" of ideas. As in Halliday and Hasan's system of analysis, these words or tokens could be repetitions of the same idea, synonyms, or reference items which presuppose the same meaning. The basic assumption underlying this portion of the analysis was that repeated reference resulted from the development of ideas which constituted central themes in the texts.

In each paper, the number of cohesive ties in the three longest text chains was counted. Lexical ties and other cohesive ties produced by substitution, reference and ellipsis were included in calculating the length of the text chains. Conjunctive ties were counted separately from other types of ties, since such ties are often used to specify the relationships between more than one idea or topic. Conjunctive ties were tabulated as additive (e.g., and, and also, nor, neither, or, or else, and...not, alternatively, etc.), adversative (e.g., yet, but, however, although, nevertheless, still, etc.), causal (e.g., so, the, thus, therefore, since, because of, etc.), or temporal (e.g., next, before, then, after, subsequently, this time, etc.).

3.1.3 Errors. Transcription errors, grammatical errors, and errors related to standards of usage for written language were also counted in each paper of the sample. The errors were classified into 12 categories:

1. Unattached dependent clauses (e.g., "When he was fifteen."), and structures which lacked a complete verb (e.g., "The boy running down the hill."), were counted as fragments.
2. Comma splices were counted if a comma was used to join two or more independent clauses when there should have been two sentences separated by a period or a semi-colon (e.g., "I chose to stay because of the fact that some people are different from others, they have their own perceptions and...").
3. Comma omissions were counted when commas were

missing after initial introductory clauses, etc. (e.g., "When the boy came running we saw..."). They were not counted when they were omitted around phrases in the second clause of sentences that contained two coordinated clauses.

4. Subject-verb agreement errors were counted when subjects and verbs did not agree in number (e.g., "The fastest boy and the fastest girl runs tomorrow.").
5. Non-standard verb usage errors included items as "I seen...", "He brung...", items related to dialect differences, or items inappropriate for written texts.
6. Pronoun agreement errors were counted when pronouns did not agree in number with their referents (e.g., "The funny thing about the boy was that they like to swim.").
7. Non-standard pronoun usage errors included items such as "Sally and me went to the party...", "Us boys went to the beach. .," "He went with John and I," etc.
8. Ambiguous reference errors included instances where a pronoun could be interpreted as referring to more than one person or thing mentioned in the text (e.g., "Bobby smiled at Tim when he won."), as well as instances where no antecedent appeared in the text preceding the use of a pronoun.
9. Modification errors included misplaced modifiers (e.g., "There was a person on the bus with a brown cap."), two-way modifiers (e.g., "The mayor said in his speech his brother told the truth."), and dangling modifiers (e.g., "While correcting papers, the message came from the principal.").
10. Misuse of adjectives and adverb errors included adverbs used as adjectives (e.g., "The carpenters built the platform sturdily enough."), and double comparisons (e.g., "more happier," "most farthest," etc.).
11. Spelling errors.
12. The final category was used for tabulating errors other than those noted above. Items which appeared in this category included

errors such as double negatives, and words used in appropriate contexts.

### 3.2 Prompt-based Variables

The prompt-based analysis system was devised to specify task demands in the prompt. These variables could be examined in relation to the response based variables found in the student papers. Student papers were independently read by two investigators who decided whether or not the response matched the category. Areas of disagreement were resolved by consensus or by a third party arbiter.

The three categories of prompt-based variables were specified as being:

- A. Directly related to the explicitly stated task demands in the prompts;
- B. Elaborations of explicitly stated task demands in the prompts;
- C. Divergent, i.e., containing issues or themes not called for in the prompt or which contradicted explicitly stated task demands.

#### 3.2.1 Prompt-based Variables for Prompt 5. Prompt 5 read:

Think of a friend, real or imaginary, that you had when you were younger. Describe something you and this friend did together; try to show your reader the kind of person this friend was and why you chose this person as a friend.

The demands of this prompt, as they were expected to be reflected in the student responses, were:

<u>Prompt Demand</u>	<u>Category of Prompt Demand</u> (A=explicitly stated; B=elaboration; C=divergent)
1. Specified a friend relationship. ( <u>real</u> friend)	B
2. Event in recent past. (post elementary school)	A
3. Event in distant past. (pre secondary school)	A

<u>Prompt Demand</u>	<u>Category of Prompt Demand</u> (A=explicitly stated; B=elaboration; C=divergent)
4. Describes one event.	A
5. Describes related events.	B
6. Lists separate events.	C
7. Explicitly discusses character traits and personality.	C
8. Implicitly shows character through event.	A
9. Implicitly shows character through event. (i.e., comments on relationship between character and event)	B
10. Explicitly discusses why person was chosen.	C
11. Implicitly shows why person was chosen through event.	A
12. Explicitly shows why person through event. (i.e., comments on relationship between event and why person was chosen)	B
13. Describes a friend. (i.e., physical characteristics)	C
14. Specifies friend relationship. (imaginary friend)	B
15. Implicitly shows character traits and personality.	C

3.2.2 Prompt-based Variables for Prompt 6. Prompt 6 reads:

The school newspaper has asked students to submit suggestions about how to improve the school. Think of ONE and only one problem which you would like to see solved to make life at school better for you and others. In a letter to the school newspaper describe the problem and tell how you would like to see it solved.

The demands of this prompt, as they were expected to be reflected in the student responses, were:

<u>Prompt Demand</u>	<u>Category of Prompt Demand</u> (A=explicitly stated; B=elaboration; C=divergent)
1. Refers to request for suggestions.	B
2. Explicitly or obviously relates problem to school <u>improvement</u> .	B
3. Addresses <u>one</u> problem/issue.	A
4. Addresses <u>related</u> problem/issues.	B
5. Lists <u>separate</u> problems/issues.	C
6. Relates problem to other aspects of school life.	B
7. Explicitly states solution would make things better for <u>self</u> .	A
8. Explicitly states solution would make things better for <u>others</u> .	A
9. Explicitly states solution would make things better for <u>everyone</u> (i.e., staff and students).	B
10. Letter in <u>form</u> .	A
11. Letter in <u>fact</u> (i.e., first person).	A
12. <u>Describes</u> the problem.	A
13. Explains <u>why</u> it's a problem.	B
14. Explains why the problem is <u>significant</u> (i.e., in comparison to other problems).	B
15. Presents solution.	A

<u>Prompt Demand</u>	<u>Category of Prompt Demand</u> (A=explicitly stated; B=elaboration; C=divergent)
16. Discusses <u>effects</u> of solution.	B
17. Explains <u>worth</u> of solution (i.e., in relation to other issues)	B
18. Implicitly describes the problem.	C

#### 4.0 Data Tabulation and Analysis

Individual papers were coded on a yes/no basis for each of the items in each category. The papers were independently coded by two members of the research staff. There was 89% agreement between the two coders on the items for prompt 5, and 91% agreement on codings for prompt 6. On items where codings disagreed, the papers were discussed and a consensus was reached. On one student paper, there was disagreement on three of the items. This paper was eliminated from the analysis. Thus, the sub-sample for topic 6 was reduced to an N of 19.

When looking at the above-noted task demand items for prompts 5 and 6, it appears that some of the task demands are comparable across both prompts. For example, both prompts call for the student-writer to focus on a single subject. If the writer did in fact focus on one subject, an expectation which was in the topic demand category "A" (i.e., "describes one event" was explicitly stated in the prompt), the student response was coded "yes," i.e., as having met the explicit task demand. If, on the other hand, the student wrote about several unrelated events and did not focus on a single subject, then that item would be coded "no." However, this latter response would be coded "yes" with regard to the item "lists separate events" (an item in category "C": divergent response), since writing about several unrelated events or problems contradicts the prompt's instructions.

A student could also write about more than one event or problem, relating them to a central theme. In response to prompt 6, for example, one student argued that students be permitted to enroll in seven classes, rather than the allowable six, so that they could fit electives into their schedule. In constructing his argument, the student discussed various problems which

were related to flexibility in scheduling and time for elective classes. Another student wrote about the general problem of vandalism, but elaborated on this theme by citing several specific problems as examples and then suggesting solutions for each. In response to prompt 5, one student writer interwove comments about a soccer tryout and the publishing of a play (two separate events), in order to show two sides of her friend's character. Responses such as the above were coded (in category B) as elaborating on demands specifically called for by the prompt.

## 5.0 Results

Results were analyzed both within and across prompts. The within-prompt analysis provided information about interactions between the writer and the prompt, by suggesting which parts of the prompt were focused upon by different writers. Obviously, some aspects of a prompt are of more importance to a writer than other aspects. The within-prompt analysis is a preliminary attempt to identify implicit and explicit prompt demands which are most salient to the writer.

Traditionally, researchers who examine effects of prompts do so only in relation to score. However, some factors may hold constant across prompts. Across-prompt analyses in this study examine the following variables in an effort to explore such possible constants:

1. total number of words in the response;
2. legibility;
3. proportion of total errors to total number of words;
4. proportion of number of words in the three longest text chains to total number of words;
5. proportion of words in the longest text chain to total number of words.

Table 4 presents the results of the analyses performed on each of the above variables across both prompts.

As shown in Table 4, only proportion of total errors was not statistically significant across prompts. This indicates that students made the same number of errors regardless of which prompt they were given. The other four variables, however, were found to differ across prompts. In terms of statistical significance, students wrote more words in responding to prompt 5, as

well as writing longer text chains and including more words in their three longest chains, than did the prompt 6 respondents. The prompt 6 papers were found to be significantly more legible than the papers written in response to prompt 5.

The four statistically significant variables noted above were examined in relation to score. The results of this analysis are shown in Table 5.

Table 5 clearly shows a significant relationship between prompt 6 and the four variables analyzed. In prompt 5, no statistically significant relationship between score and any of the four factors was found. In prompt 6, however, all of the variables were found to have a statistically significant relationship to score. As score in prompt 6 increased, so did legibility rank. However, as score increased, proportions (with regard to total number of words in the response) of errors, number of words in the longest three text chains, and number of words in the longest chain decreased. These results are elaborated upon in Section 6.0.

As noted in Section 3.2, the task analysis system was constructed to include three categories of response-reflected task demands. The first category ("A") explicitly asked for information to be included in the response (e.g., "think of ONE and only one problem"). The second category ("B") implicitly asked for information to be included in the response, thus inviting elaboration by the writer. For example, "Describe something you and this friend did together" implies a past event which was a "one-shot" affair, not an ongoing activity. The third category ("C") analyzed response characteristics which included information that was not called for or was inappropriate to the prompt. For example, the student may have listed several events or problems when one was clearly asked for. These three categories were compared to score both within and across the two prompts.

Table 6 presents results of the analysis used to determine possible relationships between score and explicit, elaborated and inappropriate information found in the written responses. As noted in Table 6, there is no statistically significant relationship between score and any of the three independent measures in prompt 5. In prompt 6, however, results indicate that as score increases, so do instances of elaborated response. Further, as scores in prompt 6 increase, inappropriate response interpretations decrease. This is of special interest when it is remembered that the mean score for prompt 6 was higher than that of prompt

5, and is a possible result of prompt effects rather than writer-related factors.

The above result is paralleled by findings noted in Table 7. A t-test on the means across prompts for the three response categories shows: (1) that significantly more students responded to explicitly asked for information in prompt 6; (2) that there was no difference between prompts in number of elaborated responses; and, (3) that significantly more responses in prompt 5 fell into the inappropriate response category than in prompt 6. These results also appear to support the contention that there is an effect of a writing prompt which is manifested by differences in responses and by holistic score between these two prompts.

## 6.0 Discussion

The results support the viewpoint that effects of writing prompts are related in student responses and influence holistic scores. Since the writing assessment distributed the prompts randomly across schools and students, no difference in group mean scores should result. That is, the randomization procedure ensures equivalent abilities of students involved in the writing task, regardless of prompt. Yet, there were differences in mean holistic scores found across responses to prompts 5 and 6.

The differences in scores across prompts could be attributed to at least two factors. First, one of the prompts may be more difficult than the other. One way to examine this possibility is to see if writers can do what the prompt calls for. Such a technique was employed in this study. Second, differences in score across prompts could result because the readers (i.e., the holistic scorers) respond differently to students' writing based upon external, non-prompt related factors such as legibility and number of errors. This aspect was also explored in this study.

The analyses employed indicated that students did more of the things called for in prompt 6 than in prompt 5. Prompt 6 was also the one that had the higher group mean score. Thus, it seems appropriate to say that the more implicit and explicit demands in a prompt that are addressed, the higher the score. In short, the analysis system indicates that prompt 5's task demands were not well identified and responded to, implying that prompt 5 was more difficult.

It should not, however, come as a surprise that prompt 5 was more difficult than prompt 6 in terms of

task demands. Prompt 6 asked the students to think of a school-related problem, explain why it was a problem, and then suggest a solution. Prompt 5 asked students to think of a friend, then to remember and describe an event involving both the student-author and the friend. The event was then to be used to illustrate the friend's character traits, in addition to showing (through the event) why the person was chosen as a friend. It seems clear that the task demands for prompt 5 were more difficult than those of prompt 6, and that students could more easily meet the task demands in prompt 6. The fact that this was illustrated through the analysis system used in this study supports the use of such analyses as a method for examining prompt difficulty based upon task demands.

The finding with regard to errors and text chains, however, is somewhat surprising. It was expected that chain length (measured here in number of "chained" words per total number of words in the written product) would parallel increases in score, rather than decrease as implied by the results. This finding, however, could be due to an artifact of the length of the student's paper. For example, as response length increases, the proportion of chained words to total words decreases even though the number of words in a chain may be substantial. If a response consisted of 100 words, with a longest chain of 5 words, the proportion is .05. If the response is 125 words long, with chain length remaining constant, the proportion is lower (.04). Yet one cannot measure chaining purely on the basis of number of words in a chain since this could be a function of length. Thus, even though limitations exist, the proportion measure is useful when one wishes to look at a text as an entity, with chains as a part of the text, rather than the chain as an entity removed from the text.

It is probable, however, that elaboration of an idea within a chain takes place as number of words per chain increases. It is this elaboration, therefore, which may be related to an increase in score, but the necessary increase in length of written response which results from this elaboration also results in a lower proportion measure. This explanation could also account for the negative and significant correlation noted in Table 5.

The study also examined variables which remain constant across prompts. Results indicated that students made the same number of errors responding to each prompt. This further supports the view that students across the two prompts were of equal ability. Legibility and length of texts, however, were significantly different when

compared across topics. Prompt 6 texts were both more legible and longer than texts from prompt 5. The within prompt analyses, however, were ambiguous in these areas. Prompt 6 results showed that legibility increased as score increased. However, length of response was not significantly related to score in either topic, nor was legibility related to score in topic 5. It is indicated that within prompt variables (i.e., task demands) have more influence on score than do external variables such as legibility.

In conclusion, this study has met its goals as a pilot project. It is clear that task demands can be related to score, and that an analysis system modeled on the one employed herein can be valuable in determining how well students respond to a prompt, and in determining how difficult a prompt is. On the basis of these preliminary results, further research in the area is justified and recommended.

The refinement of procedures for the identification and analysis of effects of writing prompts is a necessary step toward devising ways to help teachers and evaluators improve the formulation of prompts for writing assessment and instruction. Given the widespread use of direct measures for evaluating writing competence, every effort should be made to control for unintended effects in the prompts used in these assessments.

### References

- Cooper, C., Cherry, R., Gerber, R., Fleischer, S., Copley, B., and Sartisky, M. Writing Abilities of Regularly-Admitted Freshmen at SUNY/Buffalo, Unpublished study, State University of New York at Buffalo, 1979.
- Freedman, S. W. Influences on the Evaluations of Students, Doctoral Dissertation, Stanford University, 1977.
- Halliday, M. A. K., and Hasan, R. Cohesion in English. London: Longman, 1976.
- Keech, C., and Blickhahn, K. Competency Testing in Writing: Beyond the Holistic Score. Bay Area Writing Project, School of Education, University of California-Berkeley, 1981.

Table 1

Means and Variances for Prompts 5 and 6

	N	$\bar{X}$	SD
Prompt 5	91	8.67	3.13
Prompt 6	76	9.46	3.26

35

Table 2

Distribution of Papers, by Prompt and Score,  
Randomly Selected for Analysis

Score	Number in Original Pool		Number Selected for Subsample	
	Prompt 5 (N = 91)	Prompt 6 (N = 76)	Prompt 5 (N = 20)	Prompt 6 (N = 20)
2	0	1	0	0
3	3	1	1	0
4	6	4	1	1
5	4	1	1	0
6	8	4	2	1
7	8	8	2	2
8	19	17	4	4
9	13	4	3	1
10	10	10	2	3
11	6	8	1	2
12	5	3	1	1
13	1	2	0	1
14	3	6	1	2
15	2	5	0	1
16	0	2	0	1
17	2	0	1	0
18	1	0	0	0

Table 3

Legibility Ranking Scale\*

<u>Legibility Rank</u>	<u>Criterion</u>
4	Clear and legible, even if not attractive. Margins do not interfere with reading.
3	Relatively hard to read. Indents too far, distracts reader.
2	Very hard to read, some words illegible. Omits left and right margins, writing extends to edges and distracts reader.
1	Illegible, some parts impossible to read. Random line-skipping and indentation which detracts from comprehension.

\* TAM Skills Test -- Writing Feature Analysis (Keech & Blickhahn, 1981).

Table 4

Analysis of Total Words, Legibility, Proportion of Total Chain Words, Proportion of Largest Chain Words and Proportion of Total Errors Across Prompts 5 and 6

Variable	N	$\bar{X}$	SD	df	t
Total Words (topic 5)	20	222.35	81.48	37	3.11*
Total Words (topic 6)	19	157.57	40.97		
Legibility (topic 5)	20	2.80	0.62	37	-2.18*
Legibility (topic 6)	19	3.16	0.38		
Proportion of Total Words in Longest 3 Chains to Total Words (topic 5)	20	0.16	0.03	37	4.03*
Proportion of Total Words in Longest 3 Chains to Total Words (topic 6)	19	0.12	0.03		
Proportion of Words in Longest Chain to Total Words (topic 5)	20	0.08	0.02	37	4.90*
Proportion of Words in Longest Chain to Total Words (topic 6)	19	0.05	0.02		
Proportion of Total Errors to Total Words (topic 5)	20	0.07	0.03	37	0.16
Proportion of Total Errors to Total Words (topic 6)	19	0.06	0.04		

\* significant at  $p < 0.05$

Table 5

Across Prompt Analysis  
(Mendall's Correlation Co-efficients)

---

	Score (Prompt 5) (N = 20)	Score (Prompt 6) (N = 19)
Legibility	0.24	0.29*
Proportion of Total Errors to Total Words	-0.14	-0.57*
Proportion of Words in Longest 3 Chains to Total Words	-0.18	-0.34 *
Proportion of Words in Longest Chain to Total Words	-0.15	-0.37*

---

\* Significant at  $p < 0.05$

85,

Table 6

Score Related to Prompt Based Explicit, Elaborated  
and Inappropriate Responses by Prompt  
(Kendall's Tau<sub>c</sub>)

	Explicit (proportion)	Elaborated (proportion)	Inappropriate (proportion)
Prompt 5 (N = 20)	-0.47	-0.50	0.18
Prompt 6 (N = 19)	0.11	0.31*	-0.34*

\* Significant at  $p < 0.05$

Table 7

T-Test Results, Means and Standard Deviations for Explicit,  
Elaborated and Inappropriate Responses Across Prompts

	Explicit			Elaborated			Inappropriate		
	$\bar{X}$	SD	t	$\bar{X}$	SD	t	$\bar{X}$	SD	t
Prompt 5 (N = 20)	0.01	0.13	-6.82*	0.01	0.01	0.17	0.01	0.01	4.23*
Prompt 6 (N = 19)	0.04	0.01		0.01	0.01		0.00	0.01	

\* Significant at  $p < 0.05$

A FIELD TEST OF EIGHT PROMPTS: SANTA  
CLARA COUNTY WRITING ASSESSMENT

Gerald Camp

1.0 The Scope of the Study

In the fall of 1979, the Bay Area Writing Project at the University of California, Berkeley, was contacted by the Santa Clara County AB 65 Secondary Consortium to develop a set of writing topics or prompts. These prompts were to become a part of the Consortium's bank of evaluation instruments upon which member districts might draw for the purpose of competency testing. Specifically, the Bay Area Writing Project was charged with the following tasks:

1. Develop seven or eight questions (prompts) for writing sample to be field tested with ninth and tenth grade students.
2. Coordinate field testing, with assistance of county, using same districts involved in math and reading samples (1977-78).
3. Develop a resource booklet for students and teachers outlining procedures for scoring each prompt, using holistic grading and providing sample questions to prepare students for the examination.
4. Conduct holistic grading of all prompts, with standard point scale applied to each.
5. Tabulate results of the field test. Include these data in booklet (#3).

In subsequent discussions, the Consortium determined that it was not financially feasible or necessary to obtain data from each participating district. Instead the Consortium decided it would be sufficient to obtain data on the performance of each prompt in typical county schools identified as low, mid-range, and high based on the performance of students in other measures of writing ability.

For the purposes of selecting the sample schools for the study, BAWP used the percentile scores of seniors on the writing skill portion of the 1977-78 California Assessment of Basic Skills. Schools whose students ranked in 1--3 percentile were designated low-range schools, those ranked 34-66 percentile were designated mid-range, and those ranked 67-99 percentile were designated high-range.

2.0 Developing the Prompts

The prompts were developed by a sub-committee of the Santa Clara County Writing Committee, a committee made up of outstanding writing teachers from Santa Clara County elementary and secondary schools and coordinated by Ms. Kate Sutherland, Assistant Director of the General Education Department of the Office of the Santa

Clara County Superintendent of Schools. (Committee members who worked on the development of the prompts are listed in Appendix A.) The BAWP consultant furnished the committee with a list of prompts from many sources, and members of the committee contributed topics they had used in their classrooms or districts. The BAWP consultant and the committee met several times during the fall and winter of 1979-80, narrowing the list of prompts, rewriting each of those finally selected several times until everyone was satisfied that each prompt was expressed as clearly and as simply as possible and would present few problems to students to whom it would be assigned. Sources for "borrowed" prompts are listed under "References," page 379 below.

Some of the work of this committee can be seen by comparing the first drafts of the prompts to the final drafts actually used in the field testing. Both lists follow:

Santa Clara County  
Writing Assessment Prompts

First Draft

1. There are probably several things you would prefer to be doing right now instead of sitting here writing an essay. Choose ONE of these things and describe it so that your reader can imagine it as well as you do and can understand why you would like to be doing it.  
  
The activity can be something you do every day, or something unusual, or even something imaginary that you may never actually do.
2. Imagine that you and three other people are trapped in one of the following situations and you cannot escape for several hours. Describe your feelings or tell a story about what you do to escape or to pass the time until you are rescued.  
  
A. an elevator      b. a well      c. a locked room  
d. a life-boat      e. a car going 80 mph
3. If you had to choose to be something other than a human being, what plant or animal or other form would you choose? In your writing, give your reader some idea of what you think it would be like to be that form, and of why you chose it.
4. If you could change places with someone else, who would it be? The person you write about could be a friend, a member of your family, or anyone else, living or dead or imaginary, drawn from past or present, from books, films, etc., or from your own imagination. In your writing give your reader some idea of what it would be like to be that person, and of why that life appeals to you.

5. Think of a personal experience that has in some way changed your life, either for better or for worse: a particular event; a person, a place you have visited; a book you have read; a film, television show, or play you have seen. Describe the experience in detail and explain fully why it was an important one for you.
6. People sometimes do things that make others angry, accidentally or intentionally. Describe a situation that has caused you to become angry. Explain, using reasons and examples, why this situation angered you, and describe how you dealt with the problem.
7. While visiting in your grandmother's house, you go upstairs into the attic. In a dusty corner you find a large trunk. Propped on a chair beside the trunk is a card, and on it in your grandmother's handwriting are the words, "To my only grandchild: Inside this trunk you will find what you have always wanted. Open the trunk and remember me."  
Describe the gift and how you will use it.
8. Think of a friend, real or imaginary, that you had when you were younger. Describe one thing you and this friend did together or might have done together. Try to show your reader the kind of person this friend was and why you chose this person as a friend.

#### *Letter Prompts*

9. You have been sent a bill by a computer charging you for two record albums you did not purchase. Write a letter explaining the problem so that someone in the billing department can correct the error.  
The bill is from: Melodie Record Club, 444 Tune Street, Nashville, Tenn. 78056  
The records charged to your account are: *Kiss Presents Opera Highlights* and *Led Zeppelin Country Hits*.  
Your name and address are: Chris Brown, 456 Main Street, Bay City, CA 92001  
Your account number is: 235-9863422

10. You are looking for an interesting way to spend the summer. You see the following ad and decide to apply.

Men-Women:

TRAIN NOW FOR THE FUTURE!!  
APPRENTICES WANTED: SUMMER TRAINING PROGRAM FOR TEENS

- Electrical Engineering/Plumbing/Carpentry
- Recreation Leadership/Wilderness Guide
- Veterinary and Zoo Caretaking
- Accounting and Computer Programming
- Motorcycle and Car Repair
- Secretarial/Receptionist
- Teacher/Public Health Assistant Trainer

June 20 through August 20, with school-year and post-graduation job placement guaranteed. COST: \$500, live-in.

Send resume and choice of program to:

TEEN APPRENTICES  
P.O. Box 1111  
Bay City, CA 97006

Write a letter applying for admission to this summer training program. Specify which part of the program you would like to join, and what interests, experiences, or personal characteristics especially qualify you. Explain whether or not you will need scholarship aid.

Pretend your name is Chris Brown and you live at 456 Main Street, Bay City, CA 92001.

11. The teachers in your school have asked students to submit suggestions about how to improve the school. Think of ONE and only one problem which you would like to see solved to make life at school better for you and others. In a letter to the faculty describe the problem you would like to see changed, give any suggestions you might have about how to bring about the change, and explain how your suggestions will affect the school. Your suggestions may be serious or humorous, but they should sound sincere.
12. You have been accused by a neighbor of vandalizing the neighbor's car with a group of friends. Your parents are prepared to make you help pay for the damage. You are innocent. Write a letter to the neighbor in which you persuade him/her of your innocence and explain what you think should be done about the matter.

361

1. Imagine that you are trapped in one of the following situations and you cannot escape for several hours. Tell what you do to escape or to pass the time until you are rescued.
  - a. an elevator
  - b. an abandoned mine
  - c. a locked room
  - d. a life boat
2. If you had to choose to be something other than a human being, what would you choose? Give your reader some idea of what it would be like to be that form, and why you chose it.
3. Think of a personal experience that has in some way changed your life: a particular event; a person; a place you have visited; a book you have read; a film, television show, or play you have seen. Describe the experience and explain why it was an important one for you.
4. While visiting in your grandmother's house you find a large trunk. Propped on a chair beside the trunk is a card, and on it in your grandmother's handwriting are the words, "To my only grandchild: Inside this trunk you will find what you have always wanted."  
Describe the gift and how you will use it.
5. Think of a friend, real or imaginary, that you had when you were younger. Describe something you and this friend did together. Try to show your reader the kind of person this friend was and why you chose this person as a friend.
6. The school newspaper has asked students to submit suggestions about how to improve the school. Think of ONE and only one problem which you would like to see solved to make life at school better for you and others. In a letter to the school newspaper describe the problem and tell how you would like to see it solved.
7. Employment agencies can be very helpful in finding you a job if they know what job you want and are qualified for. Write a letter to The ABC Employment Agency, 204 East Main Street, Bay City, CA 94606. Explain what job you would like and what your qualifications are. Explain the reasons you think you should be hired and give examples of your strengths. Include references.
8. You have been sent a bill charging you for two record albums you did not purchase. Write a letter explaining the problem so that someone in the billing department can correct the error.

The bill is from Melodie Record Club, 444 Tune Street, Nashville, Tenn. 78056.

The records charged to your account are *Kiss Presents Opera Highlights* and *Led Zepplin Country Hits*.

Your name and address are: Chris Brown, 456 Main Street, Bay City, CA 92001. Your account number is: 35-800-21.

### 3.0 Designing the Sample

To ensure an adequate number of student papers on each prompt to provide a valid sample, the BAWP consultant worked with Catharine Keech, BAWP Research Assistant and a Ph.D. candidate in education at U.C. Berkeley specializing in writing assessment. Her analysis of the necessary numbers for adequate sampling is included as Appendix B.

Based on Keech's information, BAWP presented the Consortium with three sampling plans from which to select. These plans are all based on a field test of the original twelve prompts; subsequently, the Consortium chose to test eight of the twelve.

#### *Plan A: Comprehensive District Sampling*

Will provide data on how each prompt performs in each district.

#### *Procedure*

1. Sample each school in each district to obtain reliable district data on each prompt.
2. Minimum number of papers: 12,000. Minimum number of classes necessary for reading: 396 at 30 per class.
3. Important to communicate with a responsible person at each school (English Department Chairperson, Director of Guidance, etc.) to insure a *stratified* sample (equal number of ninth and tenth graders, proportional number of high achievers, average students, low achievers).

#### *Plan B: County-wide Stratified Sampling by Ability Levels*

Will provide data on how each prompt performs county-wide with students identified as high achievers, average achievers, and low achievers and with schools rated high, average and low.

Will not provide data on how each prompt performs in each district.

#### *Procedure*

1. Using socio-economic scale and/or school averages on CTBS, categorize each school as high, average or low in overall student achievement.
2. Determine the *proportion* of schools which fall in each category. Sample must be taken from the same proportion of schools in each category.
3. For the smallest category, select one or more representative schools. Administer the writing sample in classes identified as high achievers, average achievers, and low achievers. To test all twelve prompts for each ability group, we will need twenty papers per prompt per ability group, or 240 papers per group X 3 = 720 papers.

For the mid range sample, BAWP tested eight classes divided among three schools, again selecting one school in the 34-44 percentile range, one in the 45-55 range, and one in the 56-66 range.

For the high range sample, 50 percent of the total, BAWP worked with teachers in fifteen classes divided among four schools: one in the 67-77 percentile range, one in the 78-88 range, and two in the largest category, the 89-99 percentile range.

As closely as possible, classes were equally divided between ninth and tenth grade classes or BAWP used classes of mixed ninth and tenth grades. Classes tested were either those identified as "average ability" classes or, in schools in which composition classes were tracked, tests were given in one "low ability" class, two "average" classes, and one "high" class.

Either class attendance or class size was surprisingly low on the days the writing was done; as a result, we failed to get quite as many samples as hoped for. In the smallest cell, the low range schools, we obtained from seventeen to twenty-six papers on each topic; in mid range schools, between twenty-three and thirty papers on each were obtained; in high range schools, between thirty-eight and forty-seven. Despite our being a few papers short of the goal on a few of the topics, the range of scores is such that the results are adequate for field-test purposes.

#### 4.0 Conducting the Writing Sample

In each district in which the writing sample was taken, the Consortium coordinator, Dr. Burklund, contacted a district curriculum or assessment director and explained the project. The BAWP consultant then contacted the district officer, the principal at each designated school, and the English department chair person, who selected the classes to be sampled. The teacher of each class was sent a collated packet containing an equal number of test papers on each topic. Each teacher also received the following instructions:

It is important, to provide valid results, that the sampling be done as uniformly as possible. Here is what I would like you to do:

1. Select a day during the week of April 21-25 when you can devote the whole period to the writing sample.
2. On the day before, tell your students what they will be doing on the following day. Tell them the following:
  - a. They will be participating in an experiment by the University of California to see what

4. Administer writing sample in the other two school categories and ability groups in proportion to the county-wide distribution of schools in each category. (Thus if the smallest group of schools is those identified as low achievers, we would have 720 papers representing that group. If there are twice as many "average achiever" schools, we would need 1440 papers from those schools, and so on.) The total number of papers in the sample will fall between 2160 (if each group is roughly the same size) to probably no more than 5000.

*Plan C: County-wide Stratified Sampling*

Will provide data on how each prompt performs county-wide in schools identified as high, average, or low in overall student performance.

Will not provide data on students identified as high achievers, average, or low achievers.

*Procedure*

1. Categorize schools and determine proportions as in number one and number two above.
2. For the smallest category, select one or more representative schools and eight or nine representative classes. Administer writing sample to obtain twenty papers per prompt, or a total of 240 papers.
3. Administer writing sample in the other two school categories in proportion to the county-wide distribution of schools in each category. The total number of papers in the sample will fall between 720 (if each group is roughly the same size) to probably no more than 2000.

The Consortium determined that Plan C would provide adequate data and that eight prompts should be tested.

Following this decision, BAWP determined the configuration of the sample by using the 1977-78 data on the performance of seniors in county schools on the writing skills portion of the California Assessment of Basic Skills. Using these figures BAWP calculated that of the forty-four secondary schools from which data was available, ten schools or 23 percent fall within the low range (1-33 percentile), twelve schools or 27 percent fall within the mid range (34-66 percentile), and twenty-two schools or 50 percent fall within the high range (67-99 percentile).

Based on the need, then, for twenty papers per topic from the smallest cell (low range schools) or a minimum of 160 papers, BAWP determined to sample seven classes divided among three schools. To help ensure a stratified, one school in the 1-11 percentile range, one in the 12-22 range, and one in the 23-33 range were chosen.

kinds of topics students like to write on and which kinds produce their best writing.

- b. Their names will not be on the papers, so the teachers reading the papers will not know who they are. (Readers won't even know what school the papers are from.)
  - \*c. They will put numbers on their papers so the teacher can return papers to students. (You will receive the unmarked originals to use in whatever way you see as valuable.)
  - d. Different students in the class will be writing on different topics.
  - e. They will be given thirty minutes to write the papers.
  - f. Each student should be sure to bring a pen or a dark (number 2) pencil to write with. (Writing must be dark enough to photocopy or we won't be able to return the original.)
3. On the day you take the sample, follow these steps:
- a. Be sure everyone has a pen or a number 2 pencil.
  - b. Pass out the test forms. (They are collated so that we'll get an equal number of each topic. Try to pass them out in the order they're in.)
  - c. Tell the students to fill in the following numbers on the top page:  
School Code \_\_\_\_\_ Teacher \_\_\_\_\_ Class \_\_\_\_\_  
For *Student Number*, give each student his/her number in your roll book.
  - d. Tell students also to write their *Student Number* and *Class* number on the top of page 2.
  - e. Tell students to raise hands if they have a question about their topic. Deal with those individually. (Involving the whole class in discussion of a topic only a few have would be confusing.)

---

\*It later proved unfeasible to return the originals. Several factors played a part in the decision to retain them: 1) the time involved in photocopying all papers, 2) the fact that some could not be photocopied, 3) considerations of topic security. In lieu of their originals, participating teachers were sent the list of practice topics and the scoring criteria so they might practice the holistic scoring procedure with their classes if they chose.

- f. Have a supply of extra paper available in case anyone needs it. Tell them to raise hands if they need an extra sheet. Try to see that students adding a sheet put their student number on the sheet.
- g. Tell them they have thirty minutes and to begin writing.
- h. At the end of thirty minutes, call time. Allow students to finish the sentence they are writing, then collect the papers.
- i. If time remains, feel free to permit students to discuss the experience, compare notes on topics, etc.

#### 5.0 Selecting Anchor Papers

When all test papers had been returned to the county office, the BAWP consultant met with two members of the County Writing Committee, Rosalie Chako of Saratoga Elementary District, and BAWP teacher/consultant Irvin Peckham of Morgan Hill Unified, to select anchor papers for the scoring.

Personal essay papers, topics 1-6, were first separated from business letter papers, topics 7-8. Each of the three readers quickly read through roughly forty papers each, dividing the papers into two groups: top-half papers and bottom-half papers. Scoring was then refined, each paper finally receiving a rank from one (lowest) to nine (highest). The readers exchanged papers and, without seeing the previous ranking, reranked the new group. They then exchanged a third time, so that all papers had been read and ranked by all three readers.

Papers which all three readers had ranked the same were considered anchor papers. Those all three had ranked 2, 4, 6, and 8 were analyzed and their identifying qualities were listed. These lists of qualities became the scoring guidelines (rubrics) for readers.

A sampling of business letter papers was then read by the three readers, anchor papers were selected, and scoring guidelines written.

Before the day of the reading, BAWP made photocopies of the selected anchor papers to use to train the readers.

Below are the scoring guidelines for the reading in general and for the two types of topics. Anchor papers for each prompt will be found following the data on the prompt.

Santa Clara County Writing Assessment  
Scoring Guidelines

1. Use scale of 1-9, but use even points primarily. First decision: is paper in upper half or lower half? Is paper a 2, 4, 6, or 8? Then refine scoring, using odd numbers to reward things which really stand out--language use, freshness, crispness, or to lower the score of a paper which was a struggle to read.
2. Reward the writer for what he/she does well. Score all papers which can be seen in any way as responses to the topic. A writer may argue with the topic and still receive a top score. Read quickly; do not mark the papers in any way. Do not attempt to tally errors; give each paper a rapid-impression score, then go on to the next paper.
3. Generally ignore spelling and punctuation. Add a point for exceptionally sophisticated style and grace; subtract a point for severe mechanical, grammatical, and stylistic deficiencies that block communication of meaning.
4. Do not be prejudiced by handwriting unless the paper is completely illegible.

*Essay Scoring*

What follows are general descriptions of the kinds of papers you are likely to find in each scoring range. Do not feel constrained by these guidelines, however, if you find papers which illustrate strengths or weaknesses not mentioned.

- 8 - Topic is fully developed with a beginning and a sense of completion. Paper leaves reader with an emotion if a narrative paper, or persuades reader if a persuasive paper. Writer uses precise diction and is imaginative in choice of words. Paper has a sense of sincerity and will exhibit vivid descriptions or careful reasoning.
- 6 - Paper sticks to topic and develops it to some degree with specific details or examples and with some feeling evident. Paper has a generally adequate opening and closing. Sentences are varied and generally correct, though there may be some awkward sentences. Diction is accurate but shows little flair for words.
- 4 - Somewhat thin in content but uses some specific details. Little sentence variety. Sentences are generally complete, though there are likely to be awkward sentences and comma splices. Papers will often have a weak or unimaginative opening or closing and may exhibit diction problems or grammatical errors.

- 2 - Very thin content, not many concrete details. Papers will be on the topic and will develop it in at least one paragraph of five or more sentences. Sentence structure will often be immature. Papers may have very minimal openings and closings. Often writers will omit words or word endings. This score should be used for papers that reveal serious second-language problems.
- 0 - Blank papers or papers which completely ignore the topic. (Note: Take any 0 papers to the head reader.)

#### *Business Letter Scoring*

In general, mechanical correctness will weight more heavily in scoring essays. Also to be rewarded are correct business letter form, concise, to-the-point clarity, and meeting all of the requirements of the topic.

- 8 - Illustrates correct and complete business letter form including return address, inside address, appropriate opening and closing. Complaint letter will contain all information required in the topic, and will organize the information clearly, concisely, and correctly. Application letter may show some imagination, but must include all required items, and must not ramble on aimlessly.
- 6 - Shows some knowledge of correct business letter form, though parts of form not essential to delivery and accomplishment of goal may be omitted. Do not score any paper in the upper half which would not accomplish the assigned task. Papers receiving this score will, in general, not be as polished and clear as 8 papers.
- 4 - Business letter form not followed, or followed very imperfectly. Most information required by the topic is included, but may be poorly organized or not clearly expressed.
- 2 - Very inadequate as a business communication; important items of information omitted; generally very poorly organized, with little or no sense of letter form.
- 0 - Blank or non-responsive papers. (Take to head reader.)

#### 6.0 Scoring the Papers

Kate Sutherland invited teachers from several county school districts to participate as readers in the holistic scoring of the papers. Most but not all of the readers were members of the Santa Clara County Writing Committee. (Names, schools, and districts of readers are listed in Appendix C.)

Readers met at the Office of the Santa Clara County Superintendent of Schools at 9:00 A.M. on Saturday, May 10. Readers were given copies of the scoring guide for essay scoring and anchor papers representing scores of 2, 4, 6, and 8. Each reader ranked the anchor papers according to the criteria on the scoring guide. The head reader (Gerald Camp) and the two table leaders (Rosalie Chako and Irvin Peckham) led the discussion of the anchors. Readers were then given another sample paper to score. When every reader had assigned the paper a score, all scores were tallied on the chalkboard. Readers were called upon to justify their scores, and the ensuing discussion served to bring the group closer to consensus.

Six to eight additional samples were scored and discussed in the same way. When the head reader was confident that all readers understood and were consistently applying the scoring criteria, he distributed the papers for scoring.

During the scoring, the two table leaders frequently rescored randomly selected papers from readers at their tables. If a reader seemed to be consistently scoring too high or too low, the table leader or head reader would review the scoring criteria with him or her.

After reading for an hour, readers took a brief coffee break. Following the break, the entire group again scored and discussed several sample papers to help keep everyone on-target.

All papers received at least two readings. A bundle of twenty-five papers which had been read at one table would be transferred to the other table for a second reading. Each table had a code to use in scoring so the second reader would not be influenced by the score assigned by the first reader.

Papers that received scores more than one point apart were returned to the table leader for a third reading. If, after the readings, two scores were identical, the paper was assigned that score. If the paper had three scores, the middle score was assigned.

Readers finished all personal essay papers (topics 1-6), giving each two or three readings as necessary, shortly before noon. Following a lunch break, the readers reconvened, were retrained to score business letter papers using the business letter scoring guide and anchor papers, and finished the entire reading at about 3:00 P.M.

At the completion of the scoring, readers were given a questionnaire. Questions and a summary of responses were as follows:

1. Did you feel that any of the topics produced noticeably weaker writing than others? If so, which ones? Why do you think this was so?

1. Did you feel that any of the topics produced noticeably stronger writing than others? If so, which ones? Why do you think this was so?

There was very little agreement among readers as to which of the prompts were less, which more successful. Of the prompts mentioned, only number two, "If you had to choose to be something other than a human being," received a significantly negative reaction. Seven readers felt it was a poor topic whereas only one reader listed it as among the best. Reasons given for the negative reactions generally suggested the topic was not suitable for high school students, "a worn out topic from elementary days" or that many student writers failed to give reasons for their choice of particular forms.

On the positive side, six readers felt that topic number six, the letter to the school newspaper suggesting a change, produced strong writing, but two other readers identified it as a weaker topic.

Concerning the other topics there was even less consensus. Topic one, "Imagine you are trapped..." was identified as a weak topic by two readers, a strong topic by five. Topic three, a personal experience that changed your life, received no negative notices and three positive. Topic four, grandmother's trunk, was identified as a strong topic by three readers and as a weak topic by one. Topic five, a childhood friend, was seen as weak by four and strong by five responders, topic seven, the letter of application, received one negative response, and topic eight, the complaint letter, received one positive response. A general comment on business letter prompts deserves to be quoted, as it reflects an opinion mentioned orally by several readers including the table leaders, an opinion strongly held by many teachers of writing:

I found the business form letters to be counterproductive. Rather than encouraging written expression, they seemed to stifle some students and, by the nature of the prompt, penalize any attempt to editorialize, criticize, or expand the writing in any way. I found the skill to be too structured to be a measured writing skill.

From the mixed reactions of the readers we can conclude that all topics with the possible exception of topic two were successful in producing the whole range of writing quality desired in such a sampling. It is interesting to note that the statistical data (see below, pp. 356-374) would indicate that the reaction of the readers is not borne out by the scores they assigned. Scores on topic two follow a distribution pattern much like those of the other topics. One might speculate that many ninth graders are in fact more immature than the readers would like them to be and can in fact respond well to a topic some readers judge to be better suited to younger children.

3. Are there any topics you feel should be re-worded for clarity or to make them easier for students to understand? Can you suggest re-wording?

Few responded, and the suggestions generally were vague. The most useful suggestions are that topic seven, the employment agency letter prompt should include an imaginary name and return address similar to that in topic eight. I would also suggest adding the words, "Use correct business letter form" to both seven and eight.

4. What is your general feeling about the quality of writing by the students?

Four readers felt the student writing was "average," "adequate," "reasonable," and "about what I expected." Eight people were very positive about the papers they read, indicating that, given only thirty minutes, most students had responded surprisingly well. There were only two negative responses: "Students show poor ability to organize thoughts. They also don't develop ideas well." "On the whole the comps revealed a poverty of mind." As head reader, having participated in many readings, I was personally very impressed with the excellent quality of the majority of student papers.

5. How did you feel about the method of conducting the reading? If it were to be done again, what suggestions would you make for improving the procedures?

Four people suggested improvements: (1) eliminate codes for scores and use a more accurate way of identifying who read particular papers (from a table leader); (2) have all third readings of papers done by table leaders; (3) eliminate anchor papers; and, (4) better room, better coffee. The other ten responses were entirely positive, several readers noting that they had learned techniques which would be useful to them in working with students or with other teachers.

## 7.0 Results of the Study

Results of the scoring of each of the prompts is given here. Results are given in both the number of papers receiving each score and the percentage of the total. Papers which received two scores one number apart are rounded downward for lower-half papers and upward for upper-half papers. Papers scored three times which received three different scores were assigned the middle score. Scores of 0 indicate blank papers or papers that were completely non-responsive to the prompt.

Following the statistical report on each prompt are several papers representing scores of 2, 4, 6, and 8 for that prompt.

Topic 1. Imagine that you are trapped in one of the following situations and you cannot escape for several hours. Tell what you do to escape or to pass the time until you are rescued.

- a. an elevator
- b. an abandoned mine
- c. a locked room
- d. a life-boat

Score	Low-Range Schls. (1-33 percentile)		Mid-Range Schls. (34-36 percent.)		High-Range Schls. (67-99 percent.)	
	No.	Percent	No.	Percent	No.	Percent
0	0		0		0	
1	0		0		1	2
2	2	11.	2	7	4	10
3	5	26	7	24	6	15
4	5	26	7	24	11	27
5	2	11	2	7	9	22
6	3	16	5	17	6	15
7	2	11	3	10	2	5
8	0		2	7	1	2
9	0		1	3	1	2
TOTAL	19		29		41	

*Sample Papers*

Low - Score 2

I am an elevator, trapped and I can't get out, It has been along time since I been in here and nobody has came to my rescue. But I can't wait no more I am going to try and get out, but I can not find no way out, and I am scard I don't no what do. May be I should wait intill someone comes to my recue. But I been here a long time and nobody has came, so then I find a way out, it is on the of the elevator there is like a little door there, I am going to try and get up there, but it is hard to do. Because it is to high and I cannot reach It, but I try and get up by those hand holders, then I reach the top and try to open the door but Its to hard but I'm going to keep on trying till I do it Then finally It open, and Im trying to get up there, Then when I get up there the door open and the men where standing there who fixed it and told me to get down from there and the would help me. Then when I got there were people there looking at me and asking me questen, like if I was scard and how do I fill, and I told them how I felt and I felt scard and sick, that I need to lay down so they took me to a room

so I could lay down for a couple of minutes so I could get better.

#### Lower Half - Score 4

We were trapped in a locked room for four hours. At first we thought that we would die but then Phillip suggested that we try to find something to do. We all started to look for anything we could do, then Nancy found a deck of cards. After playing for about one and a half hours everyone was bored.

Later we were getting pretty scared, we thought we would never get out. Then Phillip said we can't just stand around staring at each other because it would only make it worse. We started telling each other about what we were going to do as soon as we get out. All of us had some plans and we all started to cheer up. Us girls would have really had panicked if our boyfriends hadn't had been there, but they made us feel safe with them. They suggested that us girls try to sleep for awhile while they figure out what to do. That was the last I remember when us girls woke up we were in the other room and the guys were eating pizza. Up until this day they have never told us how they got us out but I am sure glad they did.

#### Upper Half - Score 6

If I was trapped on a life boat, I would try to paddle my way to an Island or the mainland. However if I did not have a paddle things would look bleak for awhile. I would let the waves carry me somewhere. But I am not sure I would do that. In order to soothe myself down I would listen to the radio and see if any planes would be flying over by where I am. I would say to myself while I am listening to the radio I might as well prepare myself for the worst. Put on a life jacket, keep a cool head, keep the boat as well balanced as possible to keep most of the water out. After awhile I might doze off. But I shouldn't let that happen. Something could happen. I might float out to sea. Or worse. In order to keep myself awake I would do, what else, but splashwater on my face. If a plane did not come for awhile I would float around until I find something, like a piece of wood for example, to paddle. If everything else fails, I would probably have to use my arms as a paddle. I could do 50 possibly and rest and do another 50 until I reach San Francisco Bay and get help. They probably would think it is fantastic that I had survived.

High - Score 8

It was a chilly night & the sky was clear. The water below seemed like a cold, depthless pool of death. It came so quick & so unexpected, a giant boom from nowhere. I remember hitting the freezing water wondering what was going on. After a few seconds of shock I saw a yellow bobbing object. I was so cold I found it hard to move about. When the yellow object approached me, I discovered it was a lifeboat. I climbed in and looked around for other people but all I saw was burning wreckage all around me. After a while all the flames died out. It was only me & my lifeboat. It soon became day & I was cold. I was almost blue. Everything around me was white except for the gray murky fathoms of never ending water below me. Soon boredom overtook me. I began counting the pieces of boards and other miscellaneous objects that would pass by every once and a while. I tried singing but that only made me feel lonelier. I drifted on for days. I found it almost impossible to sleep when I thought of the cold. After awhile though I forgot. I went to sleep and felt my body try to shiver off the cold. I couldn't control these strange movements I was making. Finally I stopped trying & didn't care anymore. I went to sleep and I still am.

Topic 2. If you had to choose to be something other than a human being, what would you choose? Give your reader some idea of what it would be like to be that form, and why you chose it.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	0		0		1	3
1	0		0		1	3
2	2	12	0		1	3
3	2	12	4	15	8	21
4	6	35	7	26	3	8
5	1	6	7	26	8	21
6	3	18	5	19	11	29
7	1	6	3	11	5	13
8	2	12	1	4	0	
9	0		0		0	
TOTAL	17		27		38	

Sample Papers

Low - Score 2

If I had a choose to be something other than a human being I would what to be a little rabbit. On Easter the people like to play with the rabbit. There so pretty and fluffy looking. They grow up to be very big and very pretty. They like to play around with people and with kids. That is why I would like to be a little rabbit.

I choose it because it is a beautiful animal. They like to play around. There not to missy and hardly have to clean up after the rabbit. They eat all their food you give them. Thats they I choose to be a rabbit.

Lower Half - Score 4

If I wasn't a human I'd be a quarter hourse because I would like to go to the Kentucky derby some day. I would run as fast as a could to win a race. I would never buck or kick any body off. Then maybe I would make it to the Kentucky derby.

Once I make it to the Kentucky derby it would be hard. I would be trained a couple more mouths antil I beat my on record time.

It's the day of the race the grand stands are full just to see the race. Then I would walk onto the track to let the people in the grand stands see me. Then I would go to the starting gates and the race would begir.. I would try as hard as I could to win.

Upper Half - Score 6

If I was something other than a human being, I think I would like to be a bird. If I were a bird I could fly high above the crowded cities, never stopping until I was far past them into the wilderness. It would make me feel so free and no one could make me land until I wanted to. I really appreciate the freedom birds have and could sit and watch them all day.

As a bird, I could fly high above the world and look at everything from a veiw that not many people have the chance to see it from. If I flew high enough I could see great mountains as tiny hills. and see giant lakes as only droplets of water.

The feeling of freedom and indepenance would be overwhelming. I could souer high above the clouds for almost an eturnity of time, landing only to drink and eat. I would be my own master, flying only with my thoughts as company.

Birds are one of the most beautiful and grace-ful animals on earth, swooping and diving at will, in the sky. To watch a bird makes me feel as free in thought as they are in flight.

970

So a bird I would experience many beautiful things and have the freedom to choose between them.

High - Score 8

Mother Nature. Yes, that's what I'd choose to be. I'd fill the prairies with stalks of warm grass and beautiful wild weeds and flowers. There will be orchid gardens in my world with various fruit trees in which animals can delight on. If I were to be Mother Nature the world will be rich with gifts from the soil and sky. The season will come at their intervals. There will be showers in winter, burgeoning flowers, fruits and other plants in spring. The summer warmth will deliver the harvest. The beaches dance with delight as the people splash and celebrate. The effulgence of the sun will shine the warm blue sky through my paradise and at dark the brilliant stars shall twinkle. The halo of smiling Moon shall reach the farmlands, forest, prairies and waters in my beautiful paradise. I won't tolerate with evil. It shall not dwell in this paradise of mine. People and animals will be each others friends. They will help one another. If there is any conflict between the living creatures I'll handle the situation with love and understanding, which all of the creatures share in this friendly world. The beings in my world will be immortal and can change into whatever creatures they wish to be. There will be no abuses because my people and animals shall remain innocent; aware but innocent, friendly and kind. Every living creature shall work for their food. People will plow the fields. Busy cities and industries shall not exist. There will be a fixed population. No more. As Mother Nature, my duties will be to regulate and control the seasons, beings and plants in my paradise and keep the living creatures happy and well.

Topic 3. Think of a personal experience that has in some way changed your life: a particular event; a person; a place you have visited; a book you have read; a film, television show, or play you have seen. Describe the experience and explain why it was an important one for you.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	1	5	0		1	2
1	2	11	0		1	2
2	3	16	3	10	1	2
3	5	26	8	27	5	12
4	5	26	5	17	8	19
5	1	5	7	23	11	26
6	1	5	3	10	13	30
7	1	5	3	10	1	2
8	0		1	3	2	5
9	<u>0</u>		<u>0</u>		<u>0</u>	
TOTAL:	19		30		43	

*Sample Papers*

Low - Score 2

Dear school

Nothing has really changed my life except for some funny looking plants I found one day well I took it to my friends house we looked it over for a few minutes and finally decided it had to be weed so we dried it out and then went down to

Lower Half - Score 4

My Brother has definatly changed my life. If it wasn't for him I would probably not be so mean and be able to trust people more.

All my life he has teased me and harased me and I had to watch what I said or did to keep him of my case.

He has always done better than me in Math and English but has such a poor image of himself that he is jealous of me.

I can never trust my brother because he always plays tricks on me and I don't know when he will pull the next trick on me.

He has also changed my life in a way that has helped me get along better. He has always helped me with my math and English. If it wasn't for him I wouldn't be where I am today in school.

When I was a little kid my brother would go on talking forever. When I wanted to talk he would go on and not let me. After awile I would have to scream, "It is my turn to talk now." Then my mother would say, "It is Kevins turn to talk now," and then my brother would get mad at me. I think that is why I have

trouble saying things that don't sound stupid all the time.

Upper Half - Score 6

### The Turning Point

Aside from the everyday normal routine of life, there are things that interest me. One of these "hobbies" are reading books, especially science fiction and fantasy types. I must of read over fifty books of this type a year, but there's one book that stands out among all that I've read. The books title is 'The Thomas Covenant Trilogy.' This book has a great emotional impact. Remarkably, it inscribed a big mark on how I feel and react in life toward my environment. This book made me more aware of my surrounding and made me understand loneliness.

The plot of the story is a leper who goes through a world of wonder where every item is essentially important; almost alive. The words that was arranged in such a fashion formed a description. This description got absorbed into me. I actually felt what the main character felt. I started to choose sides on who I want most to survive. The everchanging story got me involved. It gave new ideas and wonders about the normal things around me. It made me more aware of my surroundings and to appreciate things of luxury instead of taking it for granted.

This was only one thing I got from the book. The book also made me understand the feeling of loneliness. While reading and getting involved, I actually got the feeling of the main character. I knew what it was like by using the character as an example. This made me react differently toward people who are usually left out of activities.

The main thing is that books can change your whole perspective, feelings, and understanding of things in your life.

High - Score 8

### The Experience of My Life

When my ninth birthday came around, one of the gifts I received from my parents was to take ballet lessons at our apartment complex's recreation building. Although I was to only take one class every two weeks, I was very excited. My dance teacher called my mother after my first lesson saying I had potential and to please keep up my lessons. Of course, I just recently discovered this phone call. My teacher's husband was the artistic director of the local ballet company and he was holding auditions for a new dance

school that would be affiliated with this company. After the audition and my first class with him, he told my parents that I was awarded a scholarship.

Ever since that scholarship I've seriously been dancing, taking one to two lessons each day. I had never known that I had enough talent to be a ballet dancer until my scholarship; and if it hadn't occurred, I might never have continued doing what I love most--dancing.

*Topic 4.* While visiting in your grandmother's house you find a large trunk. Propped on a chair beside the trunk is a card, and on it in your grandmother's handwriting are the words, "To my only grandchild: Inside this trunk you will find what you have always wanted."

Describe the gift and how you will use it.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	1	4	1	4	1	2
1	1	4	0		2	4
2	6	25	3	11	2	4
3	3	13	2	7	9	19
4	8	33	8	28	11	23
5	2	8	5	18	6	13
6	2	8	5	18	11	23
7	1	4	1	4	3	6
8	0		3	11	3	6
9	0		0		0	
	TOTAL 24		28		48	

*Sample Papers*

Low - Score 2

Inside the trunk their was three pounds of tie weed. So we had a party the next night and smoke five OZ and drank three kegs their was two hundred people at my party. The party got busted at three O clock in the morning. The pigs serched the house and found the weed i told them it was my grandmother's and they took her away.

Lower Half - Score 4

In the trunk was a bunch of twenty's, ten's

and fives. When I was down counting all the money it came out to 4,000,000 dollars. I wondered why she left it to me in her will. I was very sad because it made me remember of all the good times.

I got up and told my mom and dad about it. They told me to put it in the bank. I thought it was the best idea but I kept 1,000,000 dollars of it.

With the 1,000,000 dollars I paid of the bills like telephone, Electric, etc. I then bought a motor home for 15,960 and a Jaguar for 16,000 for my self. I also gave 11,000 to my favorite friend. I thought we should buy a new house but my parents said "not yet."

We went to Lake Tahoe that summer so my parents could gamble some. Will they gamble I went to the arcade and spent a bundle of money. After that we took a trip around the world.

Upper Half - Score 6

Slowly I opened the trunk and was amazed at what I saw. All I could see was a bright, gleaming green light that was radiating from a small stone. Around the object was the red, silk interior of the trunk. I did not understand why my grandmother would want to give me this stone. I reached down and picked the emerald colored stone up. I felt the sudden surge of electrical power vibrating through my body. My brain began to think of incredible solutions to problems that had never been solved before. I then realized that I could solve the world's problems with the gift that my grandmother left me.

High - Score 8

"What could be inside the trunk that Grandmother would have given to me?" I asked myself.

Slowly I opened the old creaky top to the trunk. My heart was pounding with excitement as I peered into the trunk.

"Magazines," I said aloud disappointedly. There must have been at least thirty magazines covering the top half of the trunk. I started taking the magazines out hoping there would be something more valuable underneath.

After all the magazines had been removed my eyes set upon two boxes. Both were rather small and black.

I carefully reached for the box on the left and opened it. Looking back up at me were two shiny stones set in earrings. They were absolutely beautiful! In the box on the right was a huge shiny stone in a necklace. There was a little card attached to the box containing the necklace. On it read "These are real diamonds Honey. The only part missing is the

ring which Rich is giving you. Congratulations my dear!"

I heard a noise behind me and quickly turned around. There was Rich with another black box. Inside that was the matching diamond ring! My engagement ring!! I was so excited that I could barely stand up! Two things I've always wanted--diamonds and Rich, and now I had them both. I would treasure these precious things always.

"Thanks Grandmother," I wispered.

Topic 5. Think of a friend, real or imaginary, that you had when you were younger. Describe something you and this friend did together. Try to show your reader the kind of person this friend was and why you chose this person as a friend.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	0		0		0	
1	1	4	1	3	1	2
2	5	19	4	14	3	7
3	4	15	7	24	5	11
4	9	35	5	17	16	36
5	3	12	6	21	4	9
6	3	12	3	10	9	20
7	1	4	2	7	1	2
8	0		0		3	7
9	0		1	3	2	5
TOTAL	26		29		44	

*Sample Papers*

Low - Score 2

My Friend

Me and my friend use to have lots of fun. We played games like Scrabble. Boggle and etc. When the bell rang for recess. We always played, tetherball, on the monkey bars, and on the swings. At lunch we would do much the same things. He would always play jokes on people. For instance, this girl got so mad at him because at lunch he would filled up a squirt gun with milk and squirt milk at her. He hardly ever got mad at me.

In May and June there where water balloons everywhere, but there was one problem we didn't have any. So we usually went home wet. When we went to Alum Rock Park we almost couldn't find our way back, but we did and that how I'm able to write this paper.

Lower Half - Score 4

### My Friend

When I use to live in Mountain View I had a true friend named Tom. We were best friends since I can remember. When I was six my friend and I used to play like we were in a war. All the kids in the block would play, but me and Tom were the best shot. We would hide in the most awkward places, thats why we were the best shot. When I was seven we got a pool put in our backyard. When the first trucks came Tom and I were watching T.V. Tom said he heard the trucks, I was so excited I ran to the door and instead of the door I went flying through the window on the side of the door. After the pool was half completed we used to run down in the pool and back up the other side.

When we were around nine we used to get big football games going in the street. Even the big kids would play. We used to also go and play golf at sunnyvale municiple golf coarse. We got into some pretty radicle games.

Tom and I used to go and catch lizards at saint Francis High school. We cougth blue bellies and Aligator lizards mainly.

Upper Half - Score 6

I sat back to in my recliner, to relax. My mind skipped back through the years, which seemed centuries, to the gentle world of my youth. When living was sweet, no worries and no cares.

I remember the old oak tree, where dad had help Mike and I build the fortress of our dreams. Mike was quiet around people, even his own family, but when we were alone, with the robins sweet song and the squirrels nimble scampers, Mike was a bold as the most voictourous men.

He told how he would in the future, rich, powerful, and proud. Even now I can visualize him, with his sparkling blue-green eyes and hair the color of wheat just about to be reaped.

Though he lived two miles from me, we were always together. Once one summer, when the sun was but rising, we "borrowed" old Mrs. Wilsons gentle bay. She used the horse only for going into town and to church. We rode to our hidden and secret lake, at least we thought it was secret. As we arrived, to see if the fishing was good, it began to sprinkle. It was odd to feel rain in the summer, but it was a short shower, and created the most beautiful rainbow,

over the lake it sparkled and glimmered.

As the sun rose to zenith, we heard the crackle of twigs behind us. We jumped to our feet, and raced to the feeding bay. Just when we were going to ride-off, my father shouted to us "Hold up, boys its only me," with a laugh.

He was nice about it, but made us tell old Mrs. Wilson that we were deeply sorry, though we weren't.

Ohh, what a long time ago! When things were simplistic and I was so innocent and young.

High - Score 8

Rae and I were the same age. I never knew her birthday, but she always seemed to keep up with me. She was my equal when it came to height and we never fought. Rae had black hair and brown eyes; my opposite but imaginary friends always get along.

One of our favorite activities was to make ourselves a part of the latest movie we had seen. At our age they were almost all put out by Walt Disney, but our favorites were the Wizard of Oz and Willie Wonka and the Chocolate Factory.

The night after we had just seen Willie Wonka and the Chocolate Factory for the first time, we immediately began on our journey through that magnificent factory. We lived through each scene, keeping the basic format but naturally altering certain parts so that we fit in more easily. Then, of course, we made up entirely new plots to add spice to the story and make it even more enjoyable.

We eventually became the heroines, along with Charlie the original hero, after fighting the vicious knids to save the oompa loompas. Well, Mr. Wonka did help a little, hitting them with his cane and generally putting them out of commission.

In the end we became part owners of the chocolate factory to visit whenever we please. Mr. Wonka had another glass elevator built for us so that we could go anywhere we pleased in a matter of seconds. Did we ever make use of that!

Topic 6. The school newspaper has asked students to submit suggestions about how to improve the school. Think of ONE and only one problem which you would like to see solved to make life at school better for you and others. In a letter to the school newspaper describe the problem and tell how you would like to see it solved.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	1	5	0		1	3
1	1	5	0		1	3
2	3	16	1	4	2	5
3	3	16	3	11	6	15
4	6	32	5	19	10	26
5	2	11	6	22	7	18
6	3	16	3	11	5	13
7	0		5	19	3	8
8	0		4	15	4	10
9	0		0		0	
TOTAL	19		27		39	

*Sample Papers*

Low - Score 2

The way the school could approve is by having everybody being friends beacouse there are to many people in this school that are prejudice and I think that it should not be that way, everybody should be friends it doesn't matter what race you are, and also the teachers you can tell that some are prejudice beacouse the way they treat you and the way you see them treat the other pupils. And maybe that will at least stop some fights against different races and it will stop riots, and it might bring people closer to each other, like walking by and say I hate that person because shes black, It shouldn't be that way. Everybody should like each other its not good to be prejudice, you would get more freinds if you werent prejudice and I think that you would have a better life knowing that you can depend on any of your friends. And I think thats the way it should be. Its up to the people the way they want to be because its there life and if they

want to be prejudice let then its them that are not going to have very many friends when they need something. So it is good for you not to be prejudice. So be friends with everyone, you'll get more out of life to have friends.

#### Lower Half - Score 4

The one problem I think our school has, is not enough people getting in the 'spirit' of things; mainly days like, 'Toga Day' and the work parties for Spirit Week.

On days like toga day, hat day, etc. a lot of people did get into the spirit of things--but not enough.

At the work parties for 'Spirit Week,' I was surprised at the number of people that didn't show up. At least I know that is true for the sophomore class. The work parties are arranged at meetings to have it at someone's house when everyone is available. But a lot of times not enough people show up.

I think this problem can be solved by putting more stress in the idea's for spirit days and spirit work parties. It could also help by having more meetings to prepare for this. More people should help spread the word (some people may not know about it), and get involved.

#### Upper Half - Score 6

Have you ever had the problem of wanting to transfer schools & found out you couldn't for some dumb reason? Chances are that you probably havn't since only about 95% of the kids ever want to transfer, but it is something you should be concerned about.

Most people don't know that it is a law of California that your parents can choose to send you to any public school they choose. So instead you find your self calling the District Office and asking why you can't transfer or going in to see the deens or the principal.

Some people just lie about there address, but most people don't. Instead, they try and do it lawfully.

But when you do talk to someone of that sort they will tell you "I'm sorry, but, unless you have a health or mental problem, you must go to the school in the boundry of which you live." They say that they need to have these rules in order to control the population in the schools but it is just not true. They have no problem with the amount of students; they have a problem with the teachers.

So I guess what you really need to be concerned about is to know your rights & don't let people rip you off for no reason.

High - Score 8

April 22, 1980

To the editor,

I am responding to the article concerning school problems which appeared in the April 17th edition of the "P\_\_\_\_\_." I feel that the athletic program at \_\_\_\_\_ should be improved. Sports at \_\_\_\_\_ are suffering from a lack of participation by students and teachers.

The main reason why students don't participate is apathy: they just don't care. They don't care about schoolwork, athletics, or even their own bodies. Teachers don't take an active role in athletics by coaching or helping out because they feel the students don't appreciate them or it just takes too much time. These problems are unfortunate, but true. Consequently, the morale here at \_\_\_\_\_ is suffering.

I feel that if athletics were publicized and incentives were given to try out for a team, this problem could be solved. If athletics soared to a new level of participation and enthusiasm, many other school problems would be solved. Students would work hard on their schoolwork so that they could have the opportunity to participate. Drug use would be lowered because drugs and athletics don't mix. Also, school spirit would rise and teachers would be more enthusiastic about coming to school.

Sincerely yours,

John Doe

*Topic 7.* Employment agencies can be very helpful in finding you a job if they know what job you want and are qualified for. Write a letter to The ABC Employment Agency, 204 East Main Street, Bay City, CA 94606. Explain what job you would like and what your qualifications are. Explain the reasons you think you should be hired and give examples of your strengths. Include references.

0111

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	0		1	4	0	
1	0		2	8	2	4
2	7	37	2	8	8	17
3	5	26	3	12	5	11
4	3	16	6	24	13	28
5	2	11	4	16	3	6
6	2	11	4	16	12	26
7	0		3	12	1	2
8	0		0		1	2
9	0		0		2	4
TOTAL	19		25		47	

*Sample Papers*

Low - Score 2

April 22

ABC Employment Agency  
204 East Main Street  
Bay City, CA 94606

I'm really interesting about finding a job. I'm in \_\_\_\_\_ high School and after I Graduate I like to go to College. Which is four years of college. I hope if you can help out to finding my job. My major's would be Mecanical noise. If I Graduate college and get the job how much do I get paid? Do I get more money if I Graduate College? My suport is Graduating College and have a job and settle my life.

Is it difficut to have a job?  
And give me some informations of how to get in and how much do we have to work and so on.  
I can speak English but not much of.  
Because it's my second language.  
I came to this country about a four years ago. And I really enjoying living here. One thing trouble is my English is not complete and easy.  
But I can study hard and go to college.  
Send me some informations of job and  
Thank you for reading.

291

Thank you again.

Your truly,

Lower Half - Score 4

ABC Employment Agency.  
204 East Main Street  
Bay City, Cal. 94601

To whom this may concern:

I think and have dreamed about being a stewardess. I think that I would make an excellent one. I love to be with people, young and old. In high school I have taken 4 years of language. I love to fly. And I like to be on the go all the time. I have never liked just sitting around.

I would like to fly with a world wide airline so that I may find out as many cultures on all the people of the world as I can. I have had some connections with people. I started out when I was twelve years old babysitting. I babysat until I was fifteen the got little jobs like being a waitress and working in department stores. They might not be luxuries but I did it all so that I could learn more about people.

Another job I would like to do to help make it a little easier for being a flight attendant is to be a bar tender. I know a lot of people like to drink when they are on a flight halfway around the world.

Well I hope you can find me something to fill my life's dream.

Sincerely

Upper Half - Score 6

ABC Employment Agency  
204 East Main St.  
Bay City CA 94606

Jeff L  
1405 Sequoia dr.  
Gilroy CA 4/23/80

Dear Sir,

I am writing this letter to inquire about employment from your agency. I am a licensed and qualified custom car painter and body man. I would like to be employed in an established and prosperous business. The reason I ask this is because I have had nothing but bad luck with new body shops that have just started out. I think that any custom shop would be glad to hire me. I have had many hours of classroom and on the job training at Gavilon College in Gilroy. I also have a complete history of my work that I'm sure an employer

would be very interested in. I have worked for two body shops in town that I will put down for references, Bob Hogue's body shop 716 Leanesly Rd-842-5671 and Rudy's body shop 441 Columbet ave,-842-6699, If you need any more information please contact me.

Sincerely,

Jeff L \_\_\_\_\_

High - Score 8

10712 F \_\_\_\_\_ Dr.  
Cupertino, CA 95014  
April 22, 1980

The ABC Employment Agency  
204 East Main St.  
Bay City, CA 94606

Dear Sir,

My name is M \_\_\_\_\_ P \_\_\_\_\_ and I am a student at \_\_\_\_\_ High School. I am 16 years old and I have lived in \_\_\_\_\_ Cupertino for 14 years.

I am looking for a part-time job. I am interested in being a bus-boy because I have had some experience. I feel by having this experience it makes me qualified to have a position as a bus-boy. Some of my other qualifications are that I enjoy being with and around people, I am a hard worker, and I always try to do the best that I possibly can.

I am listing three people that I feel know me and how I work. These people are my past employers.

\_\_\_\_\_ Restaurant  
c/o \_\_\_\_\_  
10752 \_\_\_\_\_ Rd.  
Cupertino, Ca. 95014

Mrs. \_\_\_\_\_  
10777 \_\_\_\_\_ Dr.  
Cupertino, Ca. 94014

Mr. \_\_\_\_\_  
19735 \_\_\_\_\_ Place  
Cupertino, Ca. 95014

I hope that you will consider my application and try to find a job for me. Thank you for spending the time to read my letter and I hope to hear a response from you in the near future.

Sincerely yours,

29.5

Topic 8. You have been sent a bill charging you for two record albums you did not purchase. Write a letter explaining the problem so that someone in the billing department can correct the error.

The bill is from Melodie Record Club, 444 Tune Street, Nashville, Tenn. 98056.

The records charged to your account are *Kiss Presents Opera Highlights* and *Led Zepplin Country Hits*.

Pretend your name and address are Chris Brown, 456 Main Street, Bay City, CA 92001. Your account number is 35-800-21.

Score	Low-Range Schls.		Mid-Range Schls.		High-Range Schls.	
	No.	Percent	No.	Percent	No.	Percent
0	0		2	9	0	
1	0		1	4	1	2
2	2	12	0		3	8
3	1	6	1	4	9	22
4	7	41	5	22	12	30
5	5	29	4	17	6	15
6	1	6	3	13	0	
7	1	6	5	22	5	13
8	0		2	9	4	10
9	0		0		0	
TOTAL	17		23		40	

*Sample Papers*

Low - Score 2

Dear Sir

I am writting to you about two albulms I was charged for. The albums are Kiss Presents Opera Highlights and Led Zepplin Country Hits but I didn't purchase these albums. My accont number is 35-900-21. I hope you will take care of this problem.

Chris Brown

Lower Half - Score 4

Dear Sir,  
I recieved a bill charging me for two albums I did not buy. The two albums were Kiss Presents Opera Highlights and Led Zeppelin Country Hits. I checked my albums and did not see these two albums in my collection. The bill I recieved came from 444 Tune Street Nashville Tenn. 98056 I checked my records I have found that there has been NO checked mailed to Nashville Tennessee. The only thing I can think of that went wrong was that your computer thought that I was someone else. I would be happy to buy these two albums, but I will not pay the bill until I recieve the two albums, as it may result in another mishappen. I would appreciate some information regarding this subject. I would also like to know what went on & how this occured.

Upper Half - Score 6

456 Main Street  
Bay City, CA 92001

Melodie Record Club  
444 Tune Street  
Nashville, Tenn. 98056

Dear Sirs:

My name is Chris Brown and I am writing to you in complaint to a recent bill which was sent to me by your store. The bill claims I charged two record albums which is not true. The records were Kiss Presents Opera Highlights and Led Zeppelin Country Hits. I assume there must have been a mistake somewhere in your billing department which triggered the bill to be sent to me. I would be very appreciated if you could somehow re-track the error so the bill can be voided. My account number is 35-800-21.

Thank you very much for your cooperation and I am happy to be one of your customers.

Sincerely,

Chris Brown

395

High - Score 8

Chris Brown  
456 Main Street  
Bay City, Calif.  
92001

Melodie Record Club  
444 Tune Street  
Nashville, Tenn.  
98056

4-29-80

Dear Sirs,

I have been billed for two record albums, Kiss Presents Opera Highlights and Led Zepplin Country Hits, which I did not order or receive.

My name is Chris Brown and my account number is 35-800-21. If you check your records I'm sure that you will find the mistake and clear the problem.

Thank you for your cooperation.

Sincerely,

Chris Brown

### 8.0 Classroom Practice for Writing Competency Testing

Writing short timed essays on subjects students have had no opportunity to prepare for requires somewhat different skills from those needed for the usual school writing assignments. It seems only reasonable, therefore, that students should be given some training in these skills before being required to demonstrate their competency with them.

The best training for tests of this sort is frequent in-class practice with the kinds of topics which will be used for actual testing. Students can learn how their test papers are judged if they also learn to score each other's papers themselves using the holistic method.

It is suggested, however, that teachers be discouraged from using the field-tested prompts included in this study for classroom writing, as over-exposure will make their use in actual testing invalid for some students.

In place of the tested topics, teachers are encouraged to use the practice topics below and to create their own similar topics. These topics can all be scored according to the scoring guides on pp. 351-352 above.

Santa Clara County Writing Assessment  
Practice Topics

1. Think of someone you would like to be with right now. Describe this person and your relationship with him or her in such a way that your reader can understand your choice.
2. You are a very small girl or boy being dragged along by your mother or father in a large crowd. Describe your feelings, and tell what you see, hear, smell and feel in as much detail as you can. Help your reader imagine what it is like to be there.
3. If you could change places with someone else, who would it be? The person you write about can be living, dead, drawn from past or present, from books, movies, television, or your own imagination. Try to give your reader some idea of what it would be like to be that person, and of why that life appeals to you.
4. Most of us have objects that we treasure, not just because of their value in money, but for other reasons. Describe one or more objects which are important to you. Explain why you value them.
5. Not all inventions have been good for all humanity. Name one invention we would be better off without, and make it clear why.
6. What would you like to be able to invent or create? Indicate what need your invention or creation would fill, and describe or explain it in as much detail as possible.
7. Your school newspaper announces that \$10,000 has been given to your school to improve extra-curricular programs such as art, music, drama, sports, field trips, and anything else the school wishes to apply it to. Write a letter to the school newspaper suggesting ONE thing you think the school should do with some or all of the money. Tell clearly how the money would be spent and give reasons why you think money should be spent on the suggestion.
8. Write a friendly letter to a "pen pal" in another country, telling him or her about your preparations for some holiday. Your letter should be as "newsy" as possible because your friend probably does not know how you do things where you live.
9. You ordered an engraved pen from Quality Products, Inc., 456 Treat St., Bay City, CA 94012. The pen was to be a gift, and was to have the name Lee Jacobs engraved on it. When it arrived, it was engraved Les Jacobs. Write a letter to the company explaining the error and asking them to correct it. Pretend your name is Chris Brown and your address is 305 Main St., Richford, CA 94018. Use correct business letter form.

10. You have heard that Biggles Department Store, 885 Main Street, Bay City, CA 94012 is hiring teenagers for summer work. Write a letter to Ms. Janet Jones, Personnel Manager, applying for the job. Explain what job in the store you would like and what your qualifications are. Explain the reasons you think you should be hired and give examples of your strengths. Pretend your name is Chris Brown and your address is 208 Country Lane, Pleasant City, CA 90843. Include references and follow correct business letter form.

290

## References

The following are *essential* reading for all those engaged in the creation of competency tests in writing.

Cooper, Charles R. and Odell, Lee. *Evaluating Writing: Describing, Measuring, Judging*. Urbana, Illinois: National Council of Teachers of English, 1977.

Diederich, Paul B. *Measuring Growth in English*. Urbana, Illinois: National Council of Teachers of English, 1974.

Keech, Catharine, et al. *National Writing Project Guide to Holistic Assessment of Student Writing* (tentative title). Berkeley: National Writing Project. (In preparation; available January, 1981.)

Los Angeles County Schools. *A Common Ground for Assessing Competence in Written Expression*. Los Angeles: Division of Curriculum and Instructional Services, Office of the Los Angeles County Superintendent of Schools, 1978.

Santa Clara County Committee on Writing. *Holistic Scoring: An Approach to the Assessment of Student Writing K-12*. San Jose: Office of the Santa Clara County Superintendent of Schools, 1978.

Tanner, Bernard R. and Bernstein, Ruby S. *The California High School Proficiency Examination: Evaluating the Writing Samples*. Berkeley: University of California, Berkeley/Bay Area Writing Project, Curriculum Publication No. 1, 1977.

Some of the prompts used in this study were adapted from prompts previously used by the following Bay Area Writing Project teacher/consultants:

Kate Blickhahn, Tamalpais Union High School District,  
Tamalpais

Catharine Keech, Bay Area Writing Project  
Irvin Peckham, Live Oak High School, Morgan Hill Unified  
School District, Morgan Hill

Miriam Ylvisaker and Esther Harris, Oakland Unified School  
District, Oakland

Appendices

Appendix A . . . . .  
Members of the Santa Clara County Writing Committee  
Who Assisted in Developing the Prompts

Appendix B . . . . .  
Sample Sizes for Testing Topic Performance

Appendix C . . . . .  
Teachers Who Participated in Scoring Writing  
Sample Essays

400

## Appendix A

### Members of the Santa Clara County Writing Committee Who Assisted in Developing the Prompts

Rosalie Chako, Redwood Junior High School  
Saratoga Elementary District

Dawn Chase, Westmont High School  
Campbell Union High School District

Eleanor Hester, Blossom Hill Elementary School  
Los Gatos Elementary District

Karen Lilly, Cupertino High School  
Fremont Union High School District

James Marshall, Los Altos High School  
Mountain View-Los Altos Union  
High School District

Irvin Peckham, Live Oak High School  
Morgan Hill Unified School District

Betsy Reeves, Sunnyvale Elementary School District

Elizabeth Ruttencutter, Department of Instruction  
Milpitas Unified School District

Mary Williams, Fair Intermediate School  
Franklin-McKinley School District

Ellen Wilson, Los Gatos High School  
Mountain View-Los Altos Union  
High School District

## Appendix B

### Sample Sizes for Testing Topic Performance

Several rules of thumb apply to deciding what makes an adequate sample size. There are some statistical minimums based on the need to produce a normal distribution of scores: these would require a minimum (for liberal rather than totally rigorous procedures) of twenty cases to a cell. In the study you are undertaking, this would mean twenty samples of writing on each topic from each school you wish to generalize about. For an accurate and comprehensive *district* sampling, you would need to sample each school in the district. The twenty would be a minimum set to insure statistical significance of differences, no matter how small a particular school or district population.

The minimum, however, may not provide a *representative sample*, especially of a large school. However, careful *stratified random sampling procedures* can help insure that the twenty papers on each topic are a representative performance from a given school or district. The minimum of twenty papers per topic within each school would then insure reasonable chance of showing statistically significant differences if real differences existed between topics and their effects on students.

You have two schools in your sample which have populations of only 680 students and twelve schools with fewer than 1500 students. At the risk of gathering a slightly less representative sample, for these schools you could reduce the total number of samples to ten per topic for the smallest schools, and fifteen per topic for the schools which have 1000-1400 students. Statistically, if you are concerned with how a topic performs within a district (rather than within a school) you could actually get by with only twenty papers per topic from each *district*. The problem here, of course, is that it would be a difficult task to choose the twenty students in each district who would, in their abilities, represent proportionally the spread of abilities in the district. Given all the factors that affect student writing performance, including individual school programs, I would recommend stratified sampling within schools, using the school as a unit of measure rather than the whole district, and thus gathering twenty samples per topic from each school.

Whatever your decision, these figures must give way of course to the ultimate costs. But rather than cutting down the number of samples per topic to reduce costs, I would strongly advocate cutting the number of topics tested. You could spend a great deal of money to gather only two or three or five samples from each school or district for each topic, only to find that the spread of ability represented was totally misleading. Such a study might cause you to discard good topics which would have discriminated well among good writers in a particular school, for instance, simply because only poor writers were among the few who happened to write on that topic in the "good" school of the district.

Catharine Keech  
Research Assistant  
Bay Area Writing Project

## Appendix C

### Teachers Who Participated in Scoring Writing Sample Essays

#### Head Reader

Gerald Camp  
Bay Area Writing Project  
Fremont Union High School District

#### Table Leaders

Rosalie Chako  
Redwood Junior High School  
Saratoga Elementary School

Irvin Peckham  
Live Oak High School  
Morgan Hill Unified School District

#### Readers

Elizabeth Ruch  
Independence High School  
East Side High School District

James Marshall  
Los Altos High School  
Mountain View-Los Altos Union High School District

George Dailey  
Live Oak High School  
Morgan Hill Unified School District

E. Maxine Thompson  
Britton Middle School  
Morgan Hill Unified School District

Valerie Bolaris  
Murphy Middle School  
Morgan Hill Unified School District

Diane Wieden  
Landels Elementary School  
Mountain View School District

Mary Young Williams  
Fair Intermediate School  
Franklin-McKinley School District

Terry Bernard  
Live Oak High School  
Morgan Hill Unified School District

Greg Guerin  
Gilroy High School  
Gilroy Unified School District

A. Spain  
Live Oak High School  
Morgan Hill Unified School District

Eleanor Hester  
Blossom Hill Elementary School  
Los Gatos Elementary School District

Bert Howard  
Gilroy High School  
Gilroy Unified School District

Chapter 4

A STUDY OF THE CONSTRUCTION OF THE MEANINGS  
OF A WRITING PROMPT BY ITS AUTHORS, THE  
STUDENT WRITERS, AND THE RATERS

Sandra Murphy, Karen Carroll, Charles Kinzer,  
and Ann Robyns

Part I: Overview of Goals and Procedures of the  
Study

Karen Carroll and Sandra Murphy

Part II: The Writing Prompt: Differences in  
Expectations Among Participants in  
a Writing Test Episode

Charles Kinzer, Karen Carroll, and  
Sandra Murphy

Part III: The Writing Prompt: The Process of  
Interpretation and Performance That  
It Elicits in Student Writers

Sandra Murphy, Karen Carroll, and  
Ann Robyns

## Chapter 4: Part I

### OVERVIEW OF GOALS AND PROCEDURES OF THE STUDY

Karen Carroll and Sandra Murphy

#### Introduction

In any communication situation, one participant may misconstrue the meaning another participant intends to convey. Such misunderstanding is a problem of special concern within the context of assessment.

Assessment frequently occurs in educational situations. One such situation is found in the classroom setting. The type of assessment common to the classroom occurs in the following sequence: an oral or written question prompt is posed by the teacher; an oral or written answer is constructed by the student; and an oral or written comment on the student's answer is given by the teacher. This question-answer-comment format, (cf. Mehan, 1979) is comparable in some respects to large-scale writing assessment episodes. Multiple opportunities for miscommunication exist in both of these assessment situations. However, an important distinction exists between the two cases. While the teacher may readily correct a mismatch between his/her intended meaning and the student's response in the classroom setting, in an annual large-scale assessment episode, the opportunity for correction does not exist once the student has received the prompt and begins to write. The consequences of miscommunication in the latter setting, therefore, are more grave and lasting.

The situation reported here is a writing assessment "episode." The method of assessment used in the particular "episode" under study involved an holistic appraisal of student writing.\* The issues raised by the

\*This report is based upon research supported by the National Institute of Education under Grant No. NIE G-80-0034 to the Bay Area Writing Project, University of California, Berkeley, Leo Ruth, Writing Assessment Project Director.

\*The method of holistic assessment has been described by Mullis (1981) as a procedure in which multiple reader-evaluators concentrate upon forming an overall impression of each paper relative to the other papers they have read. The primary objective of holistic scoring is the evaluation of a student's writing relative to the performance of that student's peers. Myers

study are pertinent to methods of writing assessment in general.

A writing assessment "episode" consists of the following stages: 1) the initial development of a prompt by test-makers (including the formulation of the wording of a written version); 2) the reading of the prompt (including initial reading and subsequent re-reading at any stage); 3) the writing of compositions in response to the prompt; 4) the development of a scoring rubric and; 5) the use of that rubric in a scoring session.

There are a variety of participant groups and texts involved within a writing assessment episode: 1) prompt authors, who design a written paragraph to serve as a prompt; 2) students, who write compositions in response to that prompt; and 3) evaluators, who read and score the student compositions (See Figure 1).

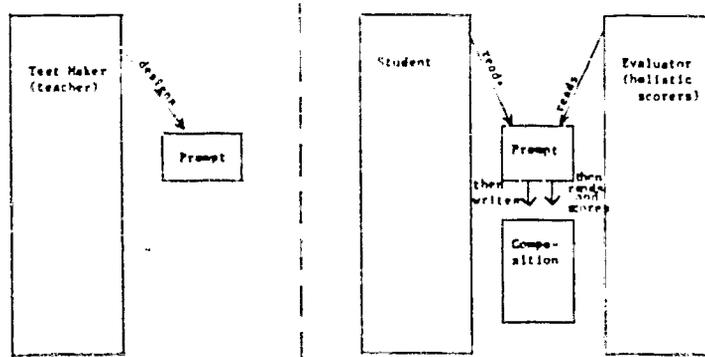


Figure 1. Participants and Texts Involved in a Writing Assessment Episode

(1930) describes the method of holistic assessment as a procedure "using an overall impression to rate a paper on a numerical scale," (p. 1).

The two texts identified in Figure 1 are the writing prompt, and the written composition produced by each student.\*

In order to determine whether participants in a typical writing assessment episode construct different interpretations of the prompt and form different interpretations of the assessment task in general, we made arrangements to follow the progress of one writing assessment episode and to record the impressions of the participants in detail at several key points along the way. These points, where the participants interact with texts in the assessment, are indicated by the arrows in Figure 1. They include the two most obviously consequential points at which miscommunication may occur: 1) the reading and interpretation of the prompt, and 2) the reading and scoring of writing samples.

The first point of participant involvement with texts in an assessment episode occurs when the ideas and structure of the prompt are formulated. Accordingly, data were collected during the process of the development of a prompt for the writing assessment under study. The second point of participant interaction with text occurred when students read the prompt prior to writing their essays. In the episode under study, interview data were collected from a subsample of 12 students before they wrote for the assessment. During the writing "test" students read the prompt and wrote an essay in response. Data collected during testing included the written compositions of all the tenth-grade students assessed at the high school, as well as the compositions written by the subsample of students who had been interviewed. The third point of participant interaction with texts occurred when evaluators read the prompt, read student responses, and scored student responses. The data collected at this point included the scores and interviews with the evaluators conducted prior to the scoring of the student essays.

Evaluation data were collected from three teachers who had developed the prompt and an additional three teachers at the high school. These teachers participated in the holistic scoring session. Another group of six teachers provided additional data by ranking the papers collected from the subsample of students. In addition to the data collected at each point of

---

\*A third text, the rubric or scoring guide, was not included in the diagram since it was not used in the analysis reported here.

prompt-participant interaction in the episode under study, questionnaire data were collected from all participants regarding assessment task and the prompt.

### The Process of Prompt Construction

The study monitored a three-member team of teachers during the development of a prompt used to assess the writing of approximately 300 tenth grade students. As they constructed the prompt, the teachers were asked to verbalize any expectations they might have about the students' written responses to the various parts of the prompt. The prompt which was eventually developed appears below:

Many different suggestions for improvement of Central High School\* have recently been made. Describe one problem or situation at Central which you feel needs correction or improvement, giving reasons for your choice and suggesting one or more solutions.

The three people involved in creating the prompt were experienced in topic development and holistic assessment. The session lasted approximately 2-1/2 hours, with several prompts being developed, considered, and ultimately rejected, in favor of the one under study. As the prompt evolved, the authors discussed their expectations of what should and might be found in responses to various parts of the prompt.

The session was tape-recorded and one member of the research team attended the session. The researcher took written notes on the issues and the potential prompts which were considered, but did not participate in the discussion except to answer procedural questions.

### Target Students

Twelve sophomores were interviewed during the normal writing assessment scheduled by the English department. The students were recommended by their teachers as average writers who would not be uncomfortable in an oral interview situation. One student was later eliminated from the sample because of problems in the tape-recording of her pre-writing interview. The final sample consisted of eleven students.

---

\*Pseudonym

## The Assessment Day Procedures

Interview. The interviews were held on the day scheduled for the writing assessment of the sophomore class. During the period preceding the student's regular English class, each of the eleven target students met one of three researchers for an individual pre-writing interview lasting approximately twenty minutes. These sessions were tape-recorded. Researchers advised students that the interviews were being conducted to investigate how people react to topics and plan their compositions.

During the interview, students were shown two prompts: a practice prompt and the prompt designed for the end-of-year sophomore class writing assessment. The practice prompt was used to familiarize the student with a "think-aloud" procedure and was given in two formats. In one format the prompt was segmented into phrases. In another format the prompt was typed in a normal block paragraph. The practice prompt asked the students to think of a friend and to describe a situation which revealed his or her character.

During each interview, the researcher presented the practice prompt in phrase-by-phrase format, asking the student to react aloud to each phrase with whatever ideas came to mind. This allowed each student to practice point by point verbalization of his or her interpretation. The student was asked to continue to "think aloud" until it was possible to settle on a specific content and focus for a composition. The student was then shown the practice prompt typed in block paragraph form, asked to re-read it, and instructed to stop anywhere in the course of this second reading to verbalize additional reactions to the task presented by the prompt.

After the practice procedure, the student was instructed to react to the second prompt, which was identified as the one to be used in the writing assessment. The assessment prompt was typed in a normal paragraph format as shown above. Initial responses were solicited by general directives such as, "What does this bring to mind?" The student was again asked to think aloud until the content and focus for the composition were chosen.

Immediately after the interview, the eleven students wrote compositions as part of the regular assessment procedure with their classmates. After writing, they returned for individual post-writing interviews, in which they were asked about variations among their initial interpretations of the prompt, plans for

composing, and the compositions that they wrote.

Questionnaire. On the day following the assessment, all of the students (N = 302) received a questionnaire designed to find out 1) how the students felt about writing in general, and 2) what they thought about the prompt. The text of the prompt was included on the questionnaire, as was a brief explanation of the scale of responses. The first 18 items probed general attitudes toward writing, and elicited a comparative evaluation of the prompt to others encountered in school. The final nine items asked the student to rate the importance of satisfying a set of task demands which could conceivably be derived from the statement of the topic. A copy of the student questionnaire is provided in Appendix A.

#### Prompt Interpretation by the Holistic Scorers

All of the student compositions produced for the assessment were scored on a Saturday morning one and a half weeks after the day of the assessment.

Six teachers served as holistic scorers. Three of the six scorers had been involved in the prompt development session, while three were "naive" in the sense that they had not participated in the formulation of the prompt.

Pre-Scoring Interviews. Interviews on the subject of topic-based expectations were conducted with the three naive scorers prior to the actual holistic scoring session. The procedure for these interviews was as follows: First, the teacher read the assessment prompt. Second, the interviewer asked the teacher to explain what the prompt required the student to do. The teacher was also requested to identify particular strategies a student might need to use in order to write a good composition in response to the prompt.

Choosing Anchor Papers. A set of 25 student compositions was chosen at random to be read and scored by all six of the scorers. Each teacher kept a list of the scores assigned to each paper. After reading the set of papers, the group tallied the scores received by each composition.

The next step in the assessment process was to select anchor papers, i.e., papers to exemplify each of the five scoring levels from 1 (lowest) to 5 (highest). The goal was to find anchors which represented a strong consensus in the scoring decisions. For example, papers chosen to exemplify a score of "5" had been placed in that category by all six scorers. In other

cases, more discussion was necessary before the teachers reached a consensus on anchor papers to represent a particular score. In addition, the tally revealed that none of this first set of papers had earned a score of "1", so it was not initially possible to identify anchor papers for the lowest category.

Development of a Rubric. Using the anchor papers as guides, the teachers constructed a rubric which described a prototypical paper for each of the scoring levels. At each level, a characteristic paper was described in terms of content, focus/organization, mechanics, and voice/style. The descriptions from the original version of the rubric are summarized in Figure 2.

Figure 2  
ORIGINAL RUBRIC

Scale	5	4	3	2	1
CONTENT	May be abstract but with specifics  All 3 parts of prompt	Weaker 5 — thinner  Can be 2 problems but announced	Either: thin development or redundant or unannounced — shift or only 2 of 3 parts	Goes nowhere Redundant Empty listing Vague	
FOCUS/ ORGANIZATION	Single focus  Unity and coherence (but may be in part...)	Single focus  Unity and coherence	Less coherent  Less focused— could be any school	Goes nowhere  Empty listing Vague	
MECHANICS	Very few — none to impede comprehension or pleasure of reading	Some slips Spelling but still not impeding	Noticeable errors that impede comprehension and annoy	Substantial fundamental errors	
VOICE/STYLE	Personal and mature  Involved with subject	Less mature diction  Less distinct voice	Impersonal and vague  Overly conversational	Unrecognizable	

Because none of the papers in the initial scoring had earned a "1", the teachers revised the rubric and the anchoring scores to differentiate the two lowest categories on the rubric. The revised rubric is presented in Figure 3.

Figure 3  
REVISED RUBRIC

Scale	5	4	3	2	1
CONTENT	May be abstract but with specifics  All 3 parts of prompt	Weaker 3—thinner  Can be 2 problems but announced	Either: thin development or redundant or unannounced shift or only 2 of 3 parts	Addresses subject Redundant, vague, empty	Does nowhere May have minimal content: 1 topic sentence 1 sentence 1 sentence 1 sentence To brief to be assessed
FOCUS/ ORGANIZATION	Single focus Unity and coherence (but may be in part...)	Single focus Unity and coherence	Less coherent Less focused—could be any school	A listing	Simplistic or random
MECHANICS	Very few—none to impede comprehension or pleasure of reading	Comma splice Spelling but still not impeding	Noticeable errors that impede comprehension and annoy	Substantial fundamental errors	Very seriously flawed
VOICE/STYLE	Personal and mature Involved with subject	Less mature diction Less distinct voice	Impersonal and vague Overly conversational	Unrecognizable	

The Scoring. Following established procedure (Myers, 1930), each composition was read and scored by two people. A five-point scale based on the rubric was used to classify each of the compositions. Two readers independently scored each composition. The scores were then combined to yield an overall score for each paper. When the two scores showed a difference of more than one point, a reconciliation by a third reader was used to resolve the discrepancy. The scoring session was tape-recorded.

The Ranking of the Target Student Papers. To provide additional data, a second direct evaluation of the target students' compositions was collected. Six experienced composition teachers were invited to perform and justify a ranking of these eleven papers. The teachers, who were all high school English instructors, were also experienced participants in holistic

assessments in their school districts. Each teacher was scheduled for an individual session with one of two researchers on the project staff. The session included a ranking of the papers followed by an interview in which the teachers were asked to discuss the criteria by which they assigned a rank for each paper. The interviews were conducted after the compositions had been ranked in order to avoid influencing the teachers' choice of criteria.

The teachers were asked to make written notes on their criteria and reactions as they went through the papers. They were able to refer to these notes during their interviews.

The first teacher in this group worked with a set of papers which included the composition written by the student who was later dropped from the sample; therefore, the data provided by this teacher was not included in our analysis.

The interview was designed to elicit the teachers' conception of what constituted a well-written composition in response to the prompt. It was also aimed at obtaining the teachers' perceptions of 1) their own criteria for evaluation of those compositions, 2) their procedures for applying those criteria, and 3) the rationale behind their criteria.

Ranking Interview Procedure. The ranking assigned to each paper was recorded. The interviewer then asked the evaluators to discuss their criteria for ranking student papers. The following list of questions was not presented directly to the teacher, but was used by the interviewer as a guideline:

1. What do you think a student would have to do in order to write a good essay in response to this topic?
2. How much weight did you give to the student's adherence to the topic/task demands in defining your criteria for evaluation?
3. How much weight did you give to other factors - e.g., spelling, basic organization, etc.?
4. In evaluating, do you tend to think of these things separately from the demands of the topic?
5. Which choices gave you the most difficulty? (i.e., choices between which papers?) and why?

The teachers were asked to describe their criteria, procedures, and rationale for ranking the compositions. Where possible, the interviewer attempted to follow the teacher's lead in the discussion, but an effort was made to ensure that the teacher also dealt with the content of all aspects of the evaluation.

Several of the rankers volunteered insights or opinions on writing instruction, evaluation of writing or related topics, which were developed through the use of non-leading interview prompts whenever possible.

## Chapter 4: Part II

### THE WRITING PROMPT: DIFFERENCES IN EXPECTATIONS AMONG PARTICIPANTS IN A WRITING TEST EPISODE

Charles Kinzer, Karen Carroll, and Sandra Murphy

#### Introduction

In order for a prompt to be meaningful it must be, first of all, comprehensible and second, capable of triggering an appropriate written output. For this to happen, certain aspects of knowledge and intention must be communicated from topic author to writer. A writer can produce successful writing in response to a given prompt when he or she can match the prompt author's expectations of what a "good" response will contain. Furthermore, if we measure success in terms of grade or score, the expectations of the evaluator must be satisfied by the writer's product. In addition, it is important to note that evaluators' expectations are triggered upon reading the prompt.

As noted above, the study monitored a three-member team during the development of a topic used to assess the writing of approximately 300 tenth grade students. As they constructed the actual prompt, the teachers were asked to verbalize any expectations they might have about the students' written responses to the various parts of that prompt. The prompt which was eventually developed appears, again, below:

Many different suggestions for improvement of Central High School have recently been made. Describe one problem or situation which you feel needs correction or improvement, giving reasons for your choice, and suggesting one or more solutions.

The questionnaire described above was given to all participants in the study. In addition, interviews were conducted with the eleven target students, with teachers who holistically scored the written products, and with five rankers who read the topic and ranked the eleven students' papers from best to worst.

The data gives us insights into the task interpretations which were "triggered" by the prompt across the three sets of participants in the assessment episode: prompt authors, student writers and holistic scorers

and/or raters. The underlying premise of this aspect of the study is that the more closely the various participants in the writing episode perceive the prompt as being similar, the more successful, in terms of score, the response.

### The Prompt Development Session

Throughout the session, the prompt authors were obviously very concerned with the implications of each word in the prompt.

Originally, the prompt asked for three improvements to the school, but this task was regarded as being "too structured". The prompt authors eventually asked for a description of one problem. In their words, they wanted to "force a choice" between possible school improvements. While students were definitely expected to select one problem or topic to write on, it was equally clear that certain kinds of topics were more "acceptable" than others. The prompt authors were somewhat worried, for example, that some responses might expound upon the merits of removing the existing principal from the school.

Although not explicitly stated in the prompt, the authors expected that the solution to the problem would be something that "money can buy". At one point, the phrase "money is no object" was written into the prompt, but this was later deleted. Although the phrase was deleted, it appeared that the topic authors expected that the "problem" selected could be "solved" through an expenditure of funds. They maintained this expectation throughout the session. References to monetary aspects of school problems and their solutions are found throughout the discussion. For example, the prompt authors felt that it might be difficult to write a rubric for scoring purposes that could apply to the subject of physical improvements to the school as well to discussions of non-material changes, such as improving school spirit. The teachers debated whether it might be difficult to equate material and non-material subjects. In effect, it was evident that students who related their problem choices and solutions to monetary factors would conform more closely to the topic authors' expectations than students who did not. Nevertheless, the authors concluded that the prompt was "wide open", as long as students could support the choice they made.

The prompt authors felt that it was not enough for the writers simply to describe the problem. Responses were expected to move beyond the description of a problem to include reasons why the problem was thought to

be a significant issue. They also indicated that the response should show that the problem was related to the writer, that the solution would benefit the writer, and that the writer felt the problem needed correction.

Key words elicited specific expectations on the part of the topic authors. For example, "situation" was added to "problem" because they felt that a problem could be "corrected," but not improved. The word "describe" was expected to invite writers to begin with strong detail and to create a strong setting. "Many" was expected to give the reader/writer an indication that there was no single, right answer to the prompt. The phrase "at Central" replaced the original "at the school" because the authors felt the latter excluded problems external to the school curriculum such as landscaping.

Thus it appears the topic authors assumed that each word in the prompt should have meaning for the writer. Although certain parts of the prompt, such as the expected strong focus on one problem, appear to have more weight than others, each segment in the prompt was intended to provide a cue to the writer for what he or she was to put in the response. More research is needed in this area of possible within-topic salience. It is probable, however, that most prompt authors would feel that any response must meet the demands which are explicit in a prompt, and that a "good" paper would demonstrate compliance with implicit demands.

### The Holistic Scorers

As noted in PART I, six scorers holistically rated the students' papers. Three of the six scorers were involved in the topic development session and three were "naive" in the sense that they were not. Interviews regarding topic-based expectations were conducted with the three naive scorers prior to the group discussion which produced the rubric upon which scoring was based. Naive scorers agreed with their counterparts who had been involved in the topic's development on almost all points. They felt strongly that it was important for the students to focus on one problem. Several scorers commented that the "structure of the paper is there in the topic." They felt the main problem student-writers would encounter was "narrowing down the focus." The following statements are representative of the naive scorers' responses upon reading the prompt:

"He (the writer) will have to declare a focus. Then he has to describe the problem and give reasons for the choice."

"They're to go from many suggestions to one problem and give reasons for the choice."

"If he manages to avoid the list and does decide on one problem . . . he'll have a good paper."

"I guess the key is finding that one problem."

"The two pitfalls may be the list . . . and the very general description of the problem."

"It has to be negative . . . they have to persuade you . . . they have to argue (that it's a problem). This calls for an argument . . . to convince someone . . . that it needs correction."

"The response needs one problem, concrete examples and specifics beyond general statements."

The above statements illustrate that the naive scorers' prompt-based expectations are extremely similar to those expressed by the prompt authors. The naive scorers' comments move beyond pure topic considerations, however, and also show a concern with style, cohesion and tone or voice. The naive scorers indicated that strong papers would be well-grounded in students' experiences at the school--that better, more credible papers would stay within personal experience.

#### Questionnaire Data on Student and Author/Rater Judgments of Task Demands

The questionnaire which students completed on the day following the assessment provided a record of their attitudes toward writing, an evaluation of the prompt, and an evaluation of task demands suggested by the prompt. Twelve of the items on the questionnaire were also presented to the teachers in the study. These items focussed on evaluation of the prompt and judgments of the task demands it might suggest. Differences among teachers' and students' perceptions of these demands are discussed below.

Responses to the items on the questionnaire were recorded on a scale of 1 to 5, with 1 signifying

agreement with a positive statement and 5 signifying agreement with a corresponding negative statement. Participants' responses are shown in Table 1 below.

Table 1

Percentage of Response by Students and Prompt Authors/Raters to Selected Questionnaire Items\*

	Agree	Neutral	Agree	
	←----->			
A. 1. The topic is more pleasing to write about than most others given in school.	31 70	37 20	31 10	The topic is less pleasing to write about than most others given in school.
2. The topic is easy to read and understand.	78 80	18 20	3 0	The topic is difficult to read and understand.
3. The topic is easy to write about.	60 60	23 20	17 20	The topic is difficult to write about.
4. I think the topic will give a good indication of my (the student's) writing ability.	21 70	37 20	42 10	I think the topic will not give a good indication of my (the student's) writing ability.
B 5. It is important to describe only one problem.	35 90	42 0	21 10	It is not important to describe only one problem.
6. It is important to explain why the subject is a problem.	92 90	13 10	3 0	It is not important to explain why the subject is a problem.
7. It is important to suggest a solution to the problem or situation.	84 80	13 20	2 0	It is not important to suggest a solution to the problem or situation.
8. It is important to suggest more than one solution to the problem.	69 10	22 20	7 70	It is not important to suggest more than one solution to the problem.
9. It is important to point out that many suggestions for improving the school have been made.	53 0	38 10	9 90	It is not important to point out that many suggestions for improving the school have been made.
10. It is important to relate the problem to everyone in the school.	71 10	22 20	7 80	It is not important to relate the problem to everyone in the school.
11. It is important to show how the problem affects the writer.	51 20	36 20	12 60	It is not important to show how the problem affects the writer.
12. It is important to explain how the solution will improve the school.	77 40	19 40	2 20	It is not important to explain how the solution will improve the school.

\*The top row of numbers in each case refers to students (N=302), the bottom row to topic authors/raters (N=10). Some rows do not total 100% due to rounding error and/or missing responses.

The table shows student-writer responses compared to the responses of the combined group of evaluators including topic authors, scorers and rankers. It is easy to see that the student writers disagreed with the other participants across many of the areas discussed.

The first four sets of figures are based on the following general statements about the prompt:

The topic is more pleasing to write about than most others given in school.

The topic is easy to read and understand.

The topic is easy to write about.

I think the topic will give a good indication of my (the students') writing ability.

It is interesting that students found the topic far less pleasing, compared to other topics given in school, than the other participants. Also, there is a definite disagreement between the students and teachers as to whether or not the writing produced by the prompt would give a good indication of writing ability. This disagreement should be cause for concern, especially since students indicated on other portions of the questionnaire that writing is an important thing to learn (84%) and that ability to write well would be important after leaving school (90%).

The combined group of evaluators, therefore, appeared to have a different idea of the relevance or quality of the prompt than did the student writers. One ranker did recognize this. She felt that the prompt would not give a good indication of writing ability, because, as she argued, topics need to be relevant to the writer and the situations discussed must be perceived as being within the writer's control. She pointed out that students may be tired of writing about school problems which they cannot directly influence and which must be solved by adults. In fact, she felt that students might resent having to spend their time providing solutions which no one would implement and that students might recognize that no one with decision-making power would read their suggestions. One of our subsample responses illustrates her point. The student recognized that she was powerless to provide a solution, saying:

"The only way we can solve this problem is...have our parents speak for us. They (school board) should listen to them because they are the ones who pay taxes."

Percentages of responses by the two groups to statements about the importance of including certain elements in compositions written in response to the prompt can be seen in Part B of Table 1.

Students and teachers appeared to agree that it was important to explain why the subject is a problem and to suggest a solution. More than 80% of both groups agreed to both of the following statements:

It is important to explain why the subject is a problem.

It is important to suggest a solution to the problem or situation.

Among the many differences found between teachers and students, one of the most dramatic was their opposing views on the importance of mentioning the suggestions which had been debated in the months preceding the assessment. Fifty-three percent (53%) of the students, but none of the teachers, agreed with the following statement:

It is important to point out that many suggestions for improving Central High School have been made.

In other words, a majority of the students thought they should reference those suggestions, while none of the teachers thought a listing of those suggestions was important.

Another difference across participant groups can be noted in responses to the statement:

It is important to describe only one problem.

In the topic development session the prompt authors had indicated they felt this was a vital element in the prompt. At one point in the topic development session, the following exchange occurred;

- a: "Do we like limiting it to one problem?  
That's what we will look for."
- b: "I do."
- c: "I do, Too."

In the questionnaire data this stress on the importance of one topic in the responses is confirmed by the combined group of evaluators. Ninety percent of the evaluators felt a focus on one problem would be a vital part of a response to this prompt. Yet these expectations were not matched by the student writers. Only 35% of the students agreed that it was important to describe only one problem.

The disagreement on the number of problems to include in the composition was consistent with the disagreement on the number of solutions to suggest. Sixty-nine percent (69%) of the students agreed with the statement: "It is important to suggest more than one solution," while only 10% of the teachers thought this was necessary. However, this result may also reflect some students' responses to the prompt segment "suggesting one or more solutions."

The two groups of participants also appeared to disagree about stressing other elements in the compositions. Many of the students emphasized elements which were not considered important by the teachers. For example, 51% of the students agreed with the statement: "It is important to relate the problem to everyone in the school," but only 40% of the teachers did.

#### Differences Between Teacher and Student Interpretations of Task Demands

The examples from the interviews and the questionnaire data indicate that teachers and students often perceived task demands differently. One major difference seemed to be that the prompt authors lent weight to all parts of the prompt: They felt that every word was important, both for the explicit meaning and the implicit requirements it conveyed. Some of the student-writers, on the other hand, focussed on only a few selected aspects of the prompt.

In the prompt development session the prompt authors appeared to feel that the first sentence provided adequate but non-prescriptive contextualization, allowing students to proceed to their selection of a specific problem. However, only one student immediately recognized that the first sentence was intended to function merely as an introduction to the prompt.

rather than as an integral part of the topic. This student said:

The way I interpreted that (the first sentence) was just as an introduction . . . I figured they were just saying this so they could tell you to describe one problem . . . I didn't think that was important . . . The most important part was (the beginning of the second sentence).

Another student demonstrated that she was aware of two possible ways of treating the first sentence. Initially, she failed to find the task interpretation favored by the teachers who developed the prompt, i.e. a focus on a single problem. She initially stated:

First I'd talk about all the suggestions that were made to improve Central, and then narrow it down to the best few and then (to) the main thing that needs improving. (Carla)

Subsequently, in her post-writing interview this same student indicated that she rejected her initial interpretation before writing:

I was thinking of doing that but then I read the topic again . . . I just thought it'd be best to start right in (on one problem).

An examination of the prompt used in this study reveals that explicit directives include the focus on only one problem, which the student feels needs to be solved, and the inclusion of reasons and one or more solutions. However, some of the teachers' expectations are not explicitly stated in the prompt. For example, (in the prompt development session) "to tell why it's a problem or situation" was the intended meaning for "giving reasons for your choice." Beyond that, many of the students felt it was important to mention that many "suggestions" had been made, while none of the teachers felt this was important. It appears, therefore, that many of the students were interpreting the task suggested by the prompt in a different way than the prompt developers intended.

The prompt or topic statement is the linking element across the three segments of a writing episode.

Therefore, it is important that the students, the test makers and test scorers agree on what the prompt is asking the student to do. Otherwise the scoring rubric may not reflect the criteria the prompt writer originally had in mind, and scoring the student papers may be further complicated if the students view the task presented by the prompt quite differently than the test makers intended.

#### Questionnaire Responses and Total Holistic Score

Kendall's coefficient of correlation was computed for total holistic score and students' responses to the pairs of items in Table 1 above. Complete texts of the student and teacher questionnaires are included in Appendix A, and B.

The measure of association was computed for both the target student group (n=11) and the larger group of students (n=302). Results are shown in Table 2.

Table 2

Kendall's Correlation Coefficients of Responses to Selected Questionnaire Items With Total Holistic Score

Item		P		P
1. The topic is more pleasing to write about than most others given in school.	.07	.035	-.06	.411
2. The topic is easy to read and understand.	-.196	.001*	-.09	.340
3. The topic is easy to write about.	-.02	.297	-.05	.410
4. I think the topic will give a good indication of my (the student's) writing ability.	.01	.478	-.12	.303
5. It is important to describe only one problem.	-.02	.289	.43	.032
6. It is important to explain why the subject is a problem.	-.16	.001*	-.29	.105
7. It is important to suggest a solution to the problem or situation.	-.13	.001*	-.28	.118
8. It is important to suggest more than one solution to the problem.	.03	.238	-.05	.417
9. It is important to point out that many suggestions for improving the school have been made.	.02	.321	.05	.417
10. It is important to relate the problem to everyone in the school.	-.09	.012	-.25	.146
11. It is important to show how the problem affects the writer.	-.02	.243	-.27	.123
12. It is important to explain how the solution will improve the school.	-.11	.002*	-.59	.006
	n=302		n=11	

\*Alpha = .05/12

In the larger group of students surveyed, significant associations were found between total holistic score (i.e. the combined scores of two evaluators) and four of the items. The results indicate that students who rated the prompt as easy to read and understand tended to receive higher holistic scores. High scores were also associated with agreement that it was important to explain why the subject chosen was a problem, and that it was important to suggest a solution to the problem. The results also indicate that students who agreed with the statement that it was important to explain how the solution/correction of the subject chosen would improve Central High tended to receive higher scores. Significant associations were not found for responses to any of the other items.

In the target group of eleven students, significant associations between questionnaire responses and holistic score were not found.

### Discussion

It is important to note in discussing these results, that although some of the measures of association are significant, they are very low. However, if one considers that many factors influence the evaluative judgments teachers make in reacting to a written composition, including factors which are not related to features of the prompt such as legibility, errors, spelling, etc., the low correlations should not be surprising. Given this qualification, the results do support the viewpoint that there is an important connection between reading and writing in a writing test situation (Ruth, 1982).

In his review of the literature on topic effects, Hoetker (1981) suggests that readability and students' reading skills are factors that should be considered in prompt design and test evaluation. These results support the notion that students with "reading" problems are also likely to have difficulty in making an appropriate response in a writing assessment. In the present study, students who found the prompt "easy to read" received higher scores.

Hoetker has also pointed out that students might have difficulty reading "between the lines" to interpret the implicit task demands suggested by a prompt. In this context, it is important to point out that two task demands which were not stated explicitly in the prompt were significantly correlated with holistic score. Agreement with the statements "It is important to explain why the subject chosen is a problem at Central," and "It is important to explain how the

solution/correction of the subject chosen will improve Central High," were significantly correlated with holistic score. Yet explaining why the subject is a problem is not precisely equivalent to either "Describe one problem," or to "giving reasons for your choice." Similarly, explaining how the subject chosen will improve Central High, is not explicitly stated as a demand of the task. Students who consider these tasks unnecessary elaborations, and/or students who interpret the task differently may be evaluated in part for their interpretation of the prompt. Students who agreed that these task demands were important were awarded higher scores.

One would expect that task demands that are stated explicitly in the prompt would be correlated with score. Thus it is not surprising that students who agreed with the statement "It is important to suggest a solution to the problem or situation" received higher scores. However, the fact that higher scores were awarded to students who recognized the importance of unstated task demands suggests that students whose task "interpretation" matched the intention of the evaluators in regard to these unstated elaborations were more likely to satisfy their expectations, and perform well in the assessment.

The question of whether or not students actually interpret the prompt in different ways is addressed more fully in Part II of this report. One should note, however, that various interpretations of the phrase "giving reasons for your choice," are possible, including giving reasons for why the problem should be considered a problem, giving reasons for why the problem was chosen above others, and/or giving reasons for why the problem was personally relevant. Given that higher scores were awarded to students who agreed that it was important to explain why the subject chosen was a problem, it appears that the first interpretation may have been favored by the evaluators. A recognition of the importance of including this explanation requires the student to go beyond the explicitly stated requirements of the prompt in order to match the evaluators' expectations.

The data presented here clearly indicate that the expectations of evaluators differ from those of students with regard to the prompt used in this assessment, and that implicit task demands may have an effect on the way students are evaluated.

THE WRITING PROMPT: THE PROCESS OF INTERPRETATION  
AND PERFORMANCE THAT IT ELICITS IN STUDENT WRITERS

Sandra Murphy, Karen Carroll, and Ann Robyns

Assumptions Underlying the Analysis

Part II of this report has dealt with differences among the prompt-based expectations of groups of individuals with regard to performance and interpretation of task. In the present discussion, we will be focusing our attention on the wording of the writing prompt, the initial (oral) interpretations of the prompt by the subsample of students described above, the ways these students interpreted the prompt, and the ways the ideas they generated were translated into written prose. The data we will be dealing with here include the oral pre-writing interviews and the written compositions collected from the sub-sample of eleven students during the writing assessment episode and their post-writing interviews.

An underlying assumption in our analysis was that the student accesses information in the process of reading the prompt which he or she then uses in composing a written response. The construction of an interpretation of the prompt is thus viewed as an integral and important part of the composing process. In the process of interpreting the prompt, the student may selectively focus on some segments, ignore others, and interpret the ones assimilated differently than do the evaluators in the assessment. Thus, the student may "construct" an interpretation of the prompt which differs from the interpretation of other participants in the assessment.

While Part II of this study revealed that students arrived at conclusions about task demands which differed from those of the prompt developers and/or evaluators, the present analysis deals with variations in the ways the eleven students constructed interpretations of the prompt.

Analysis of the Prompt

Our analysis includes an examination of the form and content of the written prompt, and examinations of the form and content of students' oral responses and written compositions. Each of these three sources of information was considered first in isolation, and then as an interactive element in a single discourse text.

Each source of data served to shed light on the others. At times, student responses focused attention on problematic features within the prompt. Conversely, certain features of the prompt, which were analyzed as being ambiguous, suggested where we might look for potential areas of confusion within student responses. By considering the interaction of the three sources of data 1) the prompt, 2) the initial interpretation/response, and 3) the written composition, we were able to identify various interpretive strategies used by the students in responding to the structure and content of the prompt within the context of this assessment situation.

### Cohesion and Contextualization

Haviland and Clark (1974, p. 3) have defined given information as: "information the speaker considers given--information he believes the listener already knows and accepts as true." They define new information as: "information the speaker considers new--information he believes the listener does not yet know." The terms given and/or "old" as they are used here refer to knowledge that is shared between "speaker" and "hearer" (in this case, prompt developers and student writers). When considered as an isolated text, there are two ways in which the following sentence could be read:

Many suggestions for improvement of Central High School have recently been made. (S1)

If the reader knows about "suggestions recently made," S1 may function as an anchor-point reference to the context of a familiar discussion. In contrast, if S1 is treated as "new" information, being introduced for the first time, it will inform the reader of an existing state of affairs of which he is as yet unaware.

If the reader treats the existence of "many different suggestions recently made" as "new" information, it is likely that the reader will expect unresolved elements of this piece of news to be qualified. Ordinarily, new information is qualified, illustrated, or expanded in the sentences that follow, to ensure that the reader is sufficiently anchored in the intended subject context. This expectation is related to the process of finding coherence between one sentence and the next.

Fillmore expresses the relationship between coherence and contextualization in the following statement:

A text can be said to be coherent if a single imagined world is compatible with all parts of the text. (1975a, p. 137)

When the reader can find no qualification, illustration or expansion of S1 by S2, he cannot readily interpret the two sentences as being coherent, and may be forced to re-examine his initial interpretation of the first sentence.

Research in the fields of cognitive processing and reading comprehension supports the view that readers expect the information in adjacent sentences to be related, and that it is easier for readers to form an integrated representation of related pieces of information if the explicit markers of cohesion in adjacent sentences are provided (Bransford and Johnson, 1973; Goldstein, 1980; Halliday and Hasan, 1976; Haviland and Clark, 1974; Hayes-Roth and Thorndyke, 1979; Kintsch, 1975; Moe, 1979; Walker and Meyer, 1980; Yekovich and Walker, 1978). The lack of direct links between the two sentences in this prompt was viewed as a potential source of difficulty for the students in the assessment.

#### Coherence and Unifying Purpose

Another criterion specified by Fillmore for coherence within a text involves "a more or less unified answer to questions about the speaker's or author's purpose," (1975a, p. 137). If the reader does not treat the information in the first sentence as new, if he treats it instead as given information, or information which refers to shared knowledge, he is forced to question the purpose the sentence would serve. The reader might ask, for example: "Is this the subject of the composition I am about to write, or is it intended to be a background for what follows?" In either case, whether the reader treats the first reference to "suggestions" as a reference to old or to new information, he is forced to search for a context and/or a unifying purpose with which to relate the two parts of the text to one another because explicit marks of cohesion between S1 and S2 do not exist.

In order to arrive at a single coherent interpretation, the reader must read selectively, applying abductive reasoning to the task. This process requires a certain degree of subtlety on the part of the reader--to know how and when to manipulate the pieces of the text to compensate for an inherent lack of coherence resulting in ambiguity. This means knowing how and when to manipulate whatever schemas, shared or personal scripts, assumptions or facts the reader has

available to him, to second guess the author's intended meaning.

We have, thus far, discussed the lack of explicit marks of cohesion between S1 and S2 of the prompt. We have argued that this constitutes an inherent source of ambiguity, resulting in two possible interpretations of the full text. In the following we will examine sentence internal presuppositions and cross-reference sentence internal information with sentence external information to determine how each may influence the other in the construction of two possible interpretations.

### Text/Semantic Analysis

The concept of presupposition as originated in logical theory appears in variously amended forms as applied by linguists in semantic analysis (Ruth, 1980, p. 129). For example, it is logically possible to distinguish between what a speaker asserts and what he presupposes, i.e., between what he says and what he takes for granted as being true. The logical presuppositions of a declarative sentence are those propositions that can be inferred in isolation from any context and that must be true if the assertion as a whole is true. Fillmore (1971) expands the concept of presupposition to include the speech communication situation, which he analyzes in terms of two levels: the presuppositional, and the illocutionary. "In every conversation, we constantly make use of both the implicit, or presuppositional, and the explicit, or illocutionary levels of communication," (Fillmore, 1971, p. 277). Thus, the construction of the meanings of given sentences necessarily requires building a structure comprised of explicit referential information plus implicit presupposed information. At this point, we should note that most theoretical discussions of presuppositional analysis pertain to idealized language users. In this case, we will apply our analysis to a real text, the Central High School writing assessment prompt, but we will describe the several idealized paths of interpretation readers might take through the text of the prompt. Our aim is to display several different potential constructions of meaning which all make sense and may account for variations in comprehension of the prompt. The text of the prompt is repeated verbatim here for reference throughout the discussion below.

Many different suggestions for improvement of Central High School have recently been made. Describe one problem or situation which you

feel needs correction or improvement, giving reasons for your choice, and suggesting one or more solutions.

In Sentence 1 (S1), the passive verb construction may shift the reader's focus of attention from the natural (cognitional<sup>#</sup>) subject (the presupposed source of "suggestions") to the natural (cognitional) object (the result of the verb process, "suggestions"), here appearing in subject position. If S1 had been cast as an active verb construction, the term "suggestions" would have appeared in its natural order in object position. Thus, there would have been represented an actor-action-goal "picture of reality" at Central High School instead of the reversed relation of goal-action and descriptive qualifiers.

When the plural marker "s" and qualifiers "many" and "different" are associated with "suggestion," the connection makes several plausible constructions of meaning or paths of interpretation possible. The following possibilities arise:

1. many different sources of opinion are engaged in the one process.
2. one source of opinion is engaged in many different processes.
3. many different sources are engaged in many different processes.

Prior experience with "suggestion-making" situations may guide the reader toward a prototypical "scenario". For example, one such scenario might involve multiple sources of opinion (representatives), offering many different suggestions in a group context (committee) on one or a few occasions (meetings).

Further information is available to the reader about the event-process(es) in the form of an adverbial modifier "recently" which circumscribes the event-process(es). In combination with the past tense marker, the modifier effectively completes or finishes the event-process(es), and places them in the "recent" past. "Suggestions have been made" orients the reader to a past event and "recently" places the event in the

---

<sup>#</sup>A term introduced by Manfred Sandmann (1954) to contrast with "grammatical" subject to avoid the confusions of "logical" and "psychological" which have been applied variously by language scholars (p. 237).

recent past, making the event a precedent for what the student writer is about to undertake.

In Sentence 2, the modifier "one," is posed as a quantitative constraint on the number of "problems or situations" to be described. It can also be interpreted, especially when S1 is acting as old information, as an indefinite pronoun. Based on the presupposition that many different opinions already exist as to what constitutes a "problem" or problematic "situation," what constitutes a "needed correction" or "improvement," and finally, what constitutes a valid "suggestion," the reader may conclude that he is required to reference "a suggestion," and describe "one" of these, i.e., "one" from a defined set of "many different suggestions."

The progress of the reader is further complicated by the presence of imperfective processes in both S1 and S2. These inherently partial, or incomplete processes are signaled by the progressive participial constructions "giving" and "suggesting." Another incomplete process is presupposed in the definition of "suggestion." The nature of a suggestion is such that it involves only a partial process within a larger process of resolution. A "suggestion" must be judged feasible before it can be instituted, and the "problem," or problematic "situation" it identifies can be resolved.

The distinction between the tasks of describing a perfective vs. an imperfective process becomes a clearer when we examine two hypothetical cases of readers: 1) those who synthesize S1 and S2; and 2) those who are unable to assimilate the lack of coherence between S1 and S2, and effectively ignore S1.

In one case, the heightened salience of repeated words or derivatives with common roots ("improvement," "suggestions," "suggesting") may lead the reader to conclude that he is expected to contribute a description of a "situation" (a state of affairs or combination of circumstances) and propose a process which will provide partial resolution ("improvement") to the "situation." In the second case, the reader may see his task as defining a problem, a definitively negative, or problematic single circumstance, and posing a definitive "solution" for "correction" of that problem.

If S2 is considered in isolation the reader may arrive at the following interpretation. The imperative "describe" presupposes the goal of the prompt task to be a description. A description is a product of personal

observation. The single instance of the imperative form indicates that the reader's observations are being elicited as the central feature of the task he is asked to perform. Presupposed here in the imperative form is the salience of these observations, supported later in S2 by the relative clause "which you feel." S2 presupposes that personally relevant observations are to form the basis for the reader's discussion. Thus, the reader is instructed to describe his observations. He is also to propose "solutions" to what he personally considers problematic, and develop or illustrate his "reasons" for why he found his choice of subject matter personally problematic, as per the segment "giving reasons for your choice."

However, if the reader interprets S2 in the context of S1, the prompt segment "giving reasons for your choice," can be interpreted as "giving reasons" why the reader chose to discuss a particular "problem or situation" above all others established in a set of precedents. In other words, the reader is reminded that he is to be held accountable for a background of shared knowledge, against which he is to justify his choice. The reader may also conclude that he is to justify his choice against a background of multiple opinions as to what constitutes a valid "suggestion," interpreting "giving reasons for your choice" as "giving reasons" for why the "problem" should be considered a "problem."

Idealized readings such as the foregoing are always exhaustive in demonstrating what can be formally interpreted from a text. Real readers, on the other hand, may choose to read selectively (to attain a unified representation of all parts of the text or merely for the sake of expediency).

Basically, what we are suggesting here, is that elements in the prompt pull the reader in different directions. The first sentence in the prompt might lead a reader to believe that he should choose from what has been posed as a defined set of suggestions, particularly if he has treated the reference to suggestions as old information. Thus, as mentioned above, the phrase "giving reasons for your choice" might be interpreted as meaning "giving reasons for why the problem was chosen above others." Other elements, on the other hand, might lead the reader to believe he has the freedom to choose any problem to discuss that is relevant to himself (cf., "which you feel") and give reasons why the problem was problematic for himself, as opposed to why the problem should be considered a problem, per se.

It has become customary for linguists to view the

text within the situation in which it was presented. In linguistic terms, the prompt represents a piece of a continuous interactive "text" with the interpretive response which is to follow. The larger context of the discourse is the writing assessment situation which involves the process of evaluation. Because of the assessment situation, readers may interpret the prompt as carrying an implicit directive: conform to the correct development of a topic in form, content, and context. Student-readers who interpret a prompt differently from the teacher-readers who evaluate their responses may be unable to conform to the evaluators' conceptions of what constitutes correct interpretation. Moreover, within the assessment situation, students are very likely to be preoccupied with the search for the right answer.

### Analysis of Oral and Written Responses to the prompt

The second body of data under investigation in our analysis is the students' oral responses to the prompt. In order to compare our analysis of the features of the prompt with the ways students responded to it during the "think-aloud" procedure conducted during the pre-writing interviews, we adopted a unit of analysis that would allow us to systematically compare the students responses with particular prompt segments. Several systems of text analysis and descriptions of discourse structure offered potential. Much of the pioneering work on the composing process has focused on narrative discourse, including descriptions of narrative structure (Mandler and Johnson, 1977, Rumelhart, 1976), and systems of text analysis which describe features associated with its production (Labov, 1972, Chafe, 1980). Because we were studying processes of interpretation and composition, systems of analysis which were based on (and derived from) actual language phenomena held the most appeal. If the point is to describe the ways individuals actually do produce meaning in the process of interpreting or responding to text, then systems of analysis that incorporate process and on line production should yield the most useful information.

Nature and Relevance of Idea Units. For this reason, the pre-writing interviews were transcribed and segmented into oral "idea units" according to criteria for oral language analysis developed by Dr. Wallace Chafe of the Linguistics Department at the University of California in Berkeley (Chafe, 1979, 1980). An idea unit is basically a chunk of words bounded by a measurable pause and/or a change in intonation, which serve as indications that the speaker is treating the word group as a conceptual unit.

Several assumptions lie behind the posited existence of such a unit. For the reader who is not familiar with Chafe's theory, they are briefly outlined here. Chafe suggests that thinking involves three major components: 1) information, 2) the self, and 3) consciousness. With regard to "information," Chafe suggests that "at any given time, a person has available a large fund of knowledge from several sources. One source is perception of the world around us, another is memory, a third is affect--the emotions, feelings and attitudes associated with what we perceive and remember" (Chafe, 1980, p. 11). Chafe also posits the existence of a "self," which he describes as an "executive" who provides the central control over what is happening " (Chafe, 1980, p. 11.). Presumably, this executive provides control over where attention is directed. Chafe also notes that this executive has a variety of needs and goals which are associated with central needs and their attainment. Consciousness, according to Chafe, possesses a central focus and a periphery, and moves in "jerks", analogous to saccadic eye movement. It is "highly limited in capacity," and it is "highly limited in the time it can dwell on one piece of information " (Chafe, 1980, p. 11.). Chafe also suggests that there are "linguistic expressions of focuses of consciousness," and these he terms "idea units." Chafe uses the idea unit as the basic unit for his system for analyzing text. Each idea unit expresses a focus of consciousness and functions to contribute to the development of knowledge which the speaker is constructing. Each idea unit represents a composite of lexical information corresponding to a simple clause, or syntactically, one verb and its associated noun phrases.

Procedure for Analyzing Oral Responses. The tape-recorded interviews were transcribed according to procedures developed by Chafe (1979, 1980). The researcher who supervised this portion of the project had previous experience in data collection and analysis using Chafe's materials and techniques in seminars and research at the University of California, Berkeley. The members of the project staff involved in preparing transcripts of the tapes were introduced to the basic concepts and procedures in a training session before transcription began.

The preparation of a transcript required that a staff member listen to the tape repeatedly, concentrating on one aspect of the language for transcription each time. An initial listening captured the basic lexical content of the interviews. The oral discourse was then segmented into "idea units" by using a combination of three types of cues: 1) intonation, 2)

hesitation, and 3) syntax. According to Chafe, intonation provides the most consistent signal of the boundaries of such a unit of discourse. Hesitations are also powerful indicators of a possible boundary between ideas, while syntactic structure is the least necessary information source for such decisions. None of the types of cues alone guarantees the existence of an idea unit, but the co-occurrence of two is a strong indication of a conceptual unit.

To prepare a transcript of each interview, research assistants were instructed to follow a six-step procedure. First, they were directed to simply record the words, i.e., the lexical content, from the tape. This step in itself involved listening repeatedly to the tape.

Second, they were directed to locate the segments of the tape containing a large number of hesitations (including both filler sounds, such as um, ah, and the lengthening of the syllables of actual words). Hesitational sounds were to be transcribed phonetically, with repeated listenings recommended 1) to sequence these correctly and 2) to verify the number of syllables or beats of filler sound in each hesitational segment. In the transcription system, a lengthened syllable was indicated by two dashes appended to the syllable.

Third, a related step involved focusing attention on the empty spaces of the discourse in order to identify the pauses and to record them in relation to the meaningful material in the transcript. Two lengths of pause time were indicated: 1) a slight break in tempo and 2) a longer pause.

Fourth, the research assistants were asked to identify three general categories of intonational contours: sentence final, clause final and question intonation. The recommended procedure was to first mark those contours which stood out most clearly and then use those contours as reference points against which to judge other intonation shifts. The distinctive falling pitch of sentence final intonation, for example, is generally more easily identified than clause-final intonation.

Fifth, a typed draft of the transcript was prepared in which each idea unit appeared on a separate line. The idea units, spoken by the interviewee, were numbered consecutively and any section where comments by the interviewer and interviewee overlapped was indicated by brackets at the points where the overlap began and ended.

marks, such as commas, dashes and periods, but discourse containing these signals can also be given different readings by different readers on separate occasions. These signals are ultimately able to serve as approximate guidelines for a reader's comprehension of the text, but they cannot record the rhythm and patterns which characterize the act of production.

Furthermore, students who are learning to compose in the written mode often have not mastered the conventions of writing. They frequently do not provide the punctuation signals which characterize mature prose, and therefore, punctuation cannot be used to guide the analysis of student writing in the same ways that it might be used with skilled adult prose. Instead, we used a combination of grammatical and semantic criteria, including recognition of shifts in point of view. The criteria are given in Figure 5.

Figure 4.

#### Criteria For Segmenting Written Discourse

In segmenting written discourse into idea units:

A. Do separate:

1. a segment ending with a period.  
This means that even clauses and phrases are treated as idea units when they end with a period.
2. a segment ending with a dash.
3. a run-on sentence into separate independent clauses.
4. a compound sentence into separate independent clauses, even if the subject is deleted in one clause.
5. a pre-posed or subordinate introductory clause,  
e.g., "Because he had to leave early, // John missed the main speech.
6. a non-restrictive clause.
7. directly quoted dialogue from the speaker tag,  
e.g., "XXXX," // said John.

Sixth, a second researcher also listened to the interview, validating the transcription and decisions about segmentation of the discourse. Any discrepancies were resolved by a third judge.

The symbols used in the transcription system are summarized in Figure 1.

Figure 5.

Symbols

- |   |       |   |
|---|-------|---|
| • Clause-final intonation<br>Contour, rising or falling<br>pitch, impression of<br>incompleteness | • •   | a break in rhythm, too<br>brief to be measured<br>(barely noticeable pause,<br>a slight break in tempo) |
| • Sentence-final intonation<br>Distinctively falling<br>pitch, impression of<br>completeness      | • • • | A pause long enough to<br>measure   |
| ? Question intonation   | --    | Lengthening of sound or<br>syllable   |

Procedure for Segmenting Written Compositions.

The written compositions were also segmented into units for analysis. The system of segmentation was based on Chafe's adaptation of his oral discourse techniques to written discourse (Chafe and Danielewicz, in preparation). However, the criteria finally used in the present analysis differed slightly from Chafe's rules for segmentation. These differences primarily involve the treatment of subordinate clauses and the phenomenon of conjoined verbs and deleted segments in sentences.

Two major problems were encountered in finalizing a set of criteria for the segmentation of the written compositions. First, existing criteria for segmenting oral discourse are heavily dependent on indications of process: pause time, hesitations, false starts, intonation, etc. (Chafe, 1980). Such signals of process are non-existent in a piece of written prose, unless a record of its production, such as that provided by a videotape, is available.

Second, the signals used in written prose to indicate where pauses or intonational contours might occur are often approximate and therefore, ambiguous. Writers indicate pauses of various lengths with punctuation

8. a rhetorical question from its answer.
9. direct representations of sounds,  
e.g., "Beep!"
10. segments containing conjoined verbs if there  
is modification after the first verb.
11. a restrictive clause which introduces new  
information.
12. a clause which conveys a shift of spatial or  
temporal orientation, or which sets up a new  
scene.
13. a clause which signals a process of recall,  
e.g., "if I remember correctly, // . . ."
14. a clause which signals personal interaction  
in the interview, i.e., which makes a text  
external reference.
15. an evaluative comment on something in the  
text.

B. Do not separate:

16. a restrictive clause which is not used to  
introduce new information.
17. a non-preposed subordinate clause, unless it  
indicates a shift of temporal or spatial  
orientation.
18. an introductory phrase, or any other phrase  
structure.
19. the subject and verb in a clause.  
This means that idea units may be embedded  
either between a subject and a verb, or  
between a subject-verb construction and its  
complement without separating that subject  
and verb into separate idea units. The sub-  
ject and verb are simply labeled with sub-  
scripts of (a) and (b) in addition to the  
number of the idea unit,  
e.g., "He<sub>a</sub>/I think<sub>b</sub> stole one of the two  
baskets of pears." : 1a/2/1b "He<sub>a</sub> stole<sub>b</sub>/I  
think<sub>c</sub>/one of the two baskets of pears." =  
1a/2/1b
20. Clausal structures which are complements to  
higher verbs,  
e.g., "I think he stole one of the two

baskets of pears." = 1 IU

"He wondered what happened to the  
pears." = 1 IU

21. an appositive from the noun to which it stands in apposition.
22. indirect speech,  
e.g., "he told me Jane was the winner." = 1 IU
23. infinitive phrases in series when they appear in the complement to the verb.
24. conjoined verbs if there is no modification after the first verb, especially if the activities are similar,  
e.g., "He pushed and shoved the door open." = 1 IU

We purposely have not called our unit of analysis for the written compositions an "idea unit", because we did not collect on-line production data in the written mode.

#### A Typology of Idea unit Functions

In order to examine the paths of interpretation taken by each student, a typology of functions was developed so that the functions of each idea unit in the interviews could be characterized. The typology was "modeled" on the typology of functions which Chafe developed for narrative (Chafe, 1980). To identify patterns of strategies that extend over several idea units, the analysis superimposed ideas from Labov's Studies of Black English Vernacular Narratives, (1972).

Chafe's typology of idea unit functions is based on the way they function as pieces of whole communications. In discussing his procedures for determining the functions of idea units, Chafe notes that the following question was asked: "In what ways does the focus of consciousness which this idea expresses contribute to the development of knowledge which, focus by focus, the speaker is constructing?" The present study required students to perform a more complicated task, than the "knowledge constructing" task Chafe posits. Our own task involved a two part process: 1) a response to a stimulus, and 2) a plan for composing. Therefore, the question which guided our analysis was: "In what ways does this idea unit contribute to the development of an interpretation of the prompt, or to the composing plan the speaker is constructing?"

In applying this procedure of questioning the function of each idea unit in the oral interviews, we found that students' comments demonstrated their awareness of their "roles" in three different "worlds." In talking with the researcher, the student was the common denominator linking the "present" interview situation, the student's "past" experience, and "future" performance in the assessment situation. Thus, the task in our study differed from that of Chafe, who characterized his own subjects as being "peripherally conscious...of at least two quite disparate worlds: that in which they found themselves sitting across from an interviewer...and the world inside the film...."\*

Within the "world" of the interview idea units served two primary functions: interaction and reference to the "think-aloud" process (i.e. the process of verbalizing a response to the prompt). It could be argued that in one sense all of the idea units within the oral interviews served the function of interaction, because we were dealing with an interactional situation. Within the world of the interview, however, certain idea units were identifiable as "overtly" interactional. In general these idea units fit into a framework where an expectation for a response from the interviewer existed, such as a question/answer format, or a framework where a student acquiesces to a directive. These included many different types of interaction, and ranged from single word comments to full clauses in response to questions.

#### Interactional.

I: . . . tell me what you think you'd have to do  
in responding to that particular topic?  
S: Ok. (Pam)  
\*\*\*\*\*  
I: Well what do you think you'll have to do?  
S: If I would . . . write about it? (Chip)

#### Think-Aloud Process.

S: . . . um . . . (sigh) hum-- . . . That's  
about all I can think of I think, (Carla)

A second world in which the students were engaged

---

\*For the reader who is unfamiliar with Chafe's analysis, we note that Chafe collected interview data based on subjects' recall of subjects, actions and events within a filmed narrative.

was the world of assessment. Comments related to assessment had a future orientation signaled by the use of the conditional. Many of these comments involved planning for the composition which the student was about to write; therefore, several of the idea units were categorized as serving the function of signaling organizational strategies or the sequence of content being planned for the essays. A wide variety of functions could potentially be identified within this general grouping.

### Planning for Composing.

- (15) S: . . . How would I go about writing it? (laughs)  
 (16) . . . Probably . . . act like,  
 (17) . . . um . . . write it more like a story. (Kim)
- (49a) S: . . . And then my third paragraph . . . um, (Kim)
- (100) S: . . . I'd re . . . read through it again,  
 (101a) see if I could make it tighter . . . in any way . . .  
 by deleting.  
 (102) . . . you know,  
 (101b) . . . words here or there. (Kim)

Still another "world" which students accessed was the world of their own personal experience/observations and background knowledge in response to the prompt. The interview served as an arena in which the student called upon previous knowledge and experience in preparing for the upcoming assessment. The students exhibited various styles of recall in the process of reviewing "past" experience as "potential" material which might be relevant to the prompt and/or the composition. Again, as in planning for the composition in the future, shifts in "worlds" were primarily signaled by shifts in verb tense, in this case, to past tense forms.

### Past Experience.

- (27) . . . and I had a ca . . . I had my schedule. (Don)
- (20) . . . well I c'n remember when they built this school,  
 (21) . . . and . . . we came over, (Bcb)

In discussing the three "worlds" in which the student operates, 1) present interview, 2) future assessment, and 3) past experience, we have introduced four of the major functions which any single idea unit

appeared to serve: Interaction, reference to process, planning for composing and reference to past (personal) experience. In addition to the functions described above, we identified two other main types of functions which individual units could serve: Reference to shared knowledge, and reference to the prompt. In the next section, these two functions will be described and discussed, and we will introduce the global function of evaluation. We should note, at this point, that like Chafe, we found that a single idea unit could serve more than one function simultaneously. Many of the idea units which served a double function were those which referred to the prompt.

Reference to the Prompt. In examining the data we discovered that many of the idea units incorporated words or groups of words from the prompt. Reference to the prompt was generally accomplished in one of three ways: 1) by the mention of the exact wording of segments of the prompt, 2) by a paraphrase of segments of the prompt, resulting in the reconstruction of the wording of the syntax of the prompt, and 3) reference to the whole prompt. In all cases, reference to the prompt often served as a point of departure which resulted in idea units serving multiple functions to suit the purpose of the student at any given instance in the interview. In referring to the prompt, students borrowed and paraphrased units of different sizes. Sometimes single words from the prompt appeared in student comments, often entire phrases or sentences were adopted.

#### Exact wording:

The student reiterates the exact wording of the prompt, or a segment of the prompt, and may incorporate it into his own statements.

- (2)           Ok (laughing, long pause) . . . 'kay, describe one problem.  
(cf. "Describe one problem")
  
- (4)           . . . Describe one problem or situation at Central which you feel needs correction or improvement.  
(Kim)  
(cf. "Describe one problem or situation at Central which you feel needs correction or improvement.")
  
- (7)           And I don't know an . . . of any . . . suggestions of . . . for improvement.  
(cf. "suggestions for improvement")

Paraphrase:

Paraphrase involved the reconstruction of the wording or the syntax of the prompt through substitution, deletion and reordering of word groups of varying size.

- (41) just describing my problem, (Kim)  
(cf. "Describe one problem")
- (23) Um . . . ck, student apathy is a problem at Central,  
(Kim) (cf. "one problem . . . at Central")
- (4) "Describe one problem or situation at Central  
which you feel needs correction or improvement."  
(5) . . . I can think of one (laughs)  
(6) . . . The lunch lines are horrible. (Katie)
- (10) . . . and "suggest one or more solutions."  
(11) . . . well . . . I would think they just  
. . . they should make a bigger cafeteria. (Katie)
- (17) And . . . about a solution to the problem  
(18) . . . I guess the . . . the solution to my problem  
would be . . . peer pressure, (Martin)
- (33) . . . there really isn't any solution to . . . the  
problem. (Jay)  
(cf. "Describe one problem . . . giving . . . one  
or more solutions.")

Reference to the prompt as a whole:

- (2) . . . um . . . I feel this . . . ur . . . is asking  
me to
- (3) . . . you know . . . tell . . . the problems of  
Central. (Carla)
- (9) It just . . . It just . . . bring Central to me. (Joe)

Frequently, references to the prompt were followed by references to personal experience, references to shared knowledge, or statements of opinion. The typical pattern involved a mention of a phrase from the prompt followed by related commentary.

Typical Pattern: Reference to Prompt/Comment

Student	Idea Unit
Katie	
Reference to prompt:	4 "Describe one problem or situation at Central which you feel needs

correction or improvement."  
 Comment: 5 . . . I can think of one (laughs)  
 6 . . . The lunch lines are horrible.  
 Reference 10 . . and "suggest one or more solutions."  
 Comment: 11 . . . well . . . I would think they  
 just . . they should make a bigger  
 cafeteria.

Student

Martin

Reference 17 And . . about a solution to the  
 to prompt: problem  
 18 . . . I guess the . . the solution to my  
 Comment: problem would be . . . peer pressure,  
 19 and you'd have to . . . get . . get  
 kids involved in wanting the school  
 to look nice.

Patterns of responses associated with reference to  
 the prompt in the interviews allowed us to examine the  
 variety of interpretations given to segments of the  
 prompt. The examples below illustrate the variety of  
 ways students responded to the first sentence (S1) of  
 the prompt in accessing (or attempting to access)  
 shared knowledge.

Reference to Shared Knowledge

(2) Um . . that comes to mind how everybody's always  
 talking about: "All we need is grass!"  
 (3) You know, "There should be a swimming pool!"  
 (4) 'Cause we were supposed to have a like uh a  
 football stadium.  
 (5) We were supposed to have . . . a swimming pool,  
 (6) bowling alleys (Alice)

\*\*\*

(70) . . . Okay . . it says "Many different suggestions  
 for improvements . . of Central . . have recently  
 been made."  
 (71) . . . so . . like . . you mean just from students.  
 (Carla)

(4) They are thinking about . . um  
 (5) . . . giving money for . . to make the school look  
 a little better.

- (6) And then--  
 (7) . . . they've . . . they've got a big grant,  
 (8) . . . state or federal or something,  
 (9) . . . to improve the curriculum  
 (10) . . . of the school. (Martin)

\*\*\*

- (7) . . . and I don't know an . . . of any suggestions  
 of . . . for improvement.  
 (8) . . . that've been made recently. (Don)

\*\*\*

- (2) . . . uh-- . . . the improvements of Central.  
 (3) . . . well one thing has been the landscaping.  
 (4) . . . it . . . like it used to be mud all over and  
 they started to put trees--,  
 (7) . . . um . . . they've painted the doors down there.  
 (8) . . . and they've kind of just changed the way  
 they've uh--  
 (9) . . . taught the . . . uh . . . students from last year.  
 (10) . . . last year was a different . . . method they  
 taught. (Jay)

Some of the students in this group indicated they were aware of suggestions that had been made (e.g., Alice). Among these students, however, the suggestions were attributed to various sources, (cf. Carla, Martin). Among the students who indicated they had some knowledge of suggestions, two indicated they had knowledge of specific suggestions which fit the temporal constraints specified in S1. These suggestions had been detailed in a school newspaper article discussing the tentative plans which school personnel were drawing up to use recently acquired grant funds for school improvement projects (e.g., Martin).

In contrast to students who claimed to have some knowledge of what the first sentence could refer to, other students indicated a lack of such knowledge (e.g., Don). Another student, who had ignored the first sentence altogether in his first interview, made the following comment when he was questioned about what he didn't like about the topic:

Just the way it's put together . . . / . . .  
 um . . . 'cause I haven't heard very many of  
 the s . . . Central High School's suggestions .  
 . . for improvement/I wouldn't know much  
 about what to write about.

The students clearly had different degrees of

familiarity with the recent debate about specific suggestions for school improvement. Their comments illustrate that various interpretations were made of the first segment of the prompt. In providing these particular examples, we are not intending to suggest that prompt authors should be held responsible for insuring that everyone being assessed share exactly the same knowledge or background information to answer a given prompt. Clearly, that would be an impossible task. It was appropriate for the authors of this prompt to assume that everyone would have something to say since it is a prompt which generally draws on knowledge of conditions at the school. However, the student responses do indicate that the first segment of the prompt was read in different ways, depending upon the nature of specific knowledge shared between the prompt authors and the eleven students. Moreover, the student responses revealed that this particular issue did cause an unanticipated range of interpretations. Our analysis suggests that the interpretive problem lies not in the fact that some students didn't know about the suggestions that had been made, but rather in the wording and the lack of specificity in the prompt itself.

In addition to the various interpretations given to sentence one, patterns of responses revealed that a variety of interpretations were given to other segments of the prompt.

Some students in this target group justified their choice of a problem in relation to other problems.

- (42) . . . that I would single that out. (attitude)
- (43) . . . and . . . and ah . . . the ah
- (44) . . . the grounds,
- (45) . . . the . . . landscaping and stuff.
- (46) . . . that's not so important. Martin (Interview)

In his written essay, Martin incorporated the same idea:

- (16) And what is landscaping
- (17) just a superficial cover that would be destroy by student with bad additude. Martin (Essay)

Other students justified their choice in terms of personal relevance:

- (24) . . . And--the reason I chose it was because in the beginning it was really bothering me when I was studying. Jay (Interview)

(201)

In Jay's written essay, the same idea is expanded to include a consequence, and a reason for that consequence.

- (9a) The reason I choose this problem is because
- (10) when I first stated high school
- (9b) my grades dropped really fast.

- (14) I could not concentrate because there was always a commotion or noise that distracted my attention from the classroom. Jay (Essay)

Other students appeared to interpret the phrase as a directive to "give reasons why a solution is needed" (i.e. why the problem should be considered a problem):

- (8) . . . Um . . . the reasons for my choice would be . . . because
- (9) . . . the school;
- (10) . . . It doesn't look good this way. Bob (Interview)

In Bob's written essay, this same idea is stated more formally:

- (6) The "unlandscaped" look the west side of the school currently has detracts from the school in many ways. Bob (Essay)

Basic Reading Ability Factors. Reading ability should also be considered. One student for example, focused on a fragment of the prompt, and treated "improvements" as the subject of the sentence in reconstructing the actual structure of the first sentence. When referring to S1, the student said:

". . . un-- . . . the improvements of Central./ . . . well one thing has been the landscaping./ . . . it . . . like it used to be mud all over and they started to put trees--./ . . . um . . . they've painted the doors down there./ . . . and they've kind of just changed the way they've un--./ . . . taught the . . . um . . . students from last year./ . . . last year was a different . . . method they taught."

Students such as this one, who have basic reading problems, are also likely to lack the ability to "read

between the lines" and to second guess the test author's unstated assumptions in relation to the prompt (Hoetker, 1932). Interpretation depends heavily on inferential ability. If a test is to present a fair and equal challenge to the writing skills of all students, ambiguity in wording must be avoided, and issues of shared knowledge should be considered.

Evaluative and Narrative Functions. As noted above, our typology was modeled on the typology of functions which Chafe developed for narrative (Chafe 1930) with Labov's work on evaluation superimposed (Labov 1972). While both Labov and Chafe incorporated evaluation in their analyses of narrative structure; Labov placed more emphasis on this global function. Chafe focused on the first three of his four main functional categories:

1. personal interaction between the speaker and his or her audience,
2. processes of recall as such,
3. the recall of a narrative as a series of states and events, and
4. evaluative comments (cf. Labov, 1972, p. 366).

Since Chafe's data involved the recall of a narrative, he expanded the third category above to include:

1. the establishment of a setting,
2. temporal orientation,
3. introducing a person,
4. characterizing a person,
5. locating a person or an object,
6. engaging a person in a durative background activity,
7. engaging a person in a punctual event, and
8. reintroducing a person.

Given that Chafe's data dealt with the recall of a narrative, his extended categorization of the function associated with the introduction of characters and their engagements in a series of states and events is clearly appropriate to his data. Narrative form and purpose revolve around characters and their activities.

Labov, in contrast, emphasizes the role of evaluation in his own analysis of narratives of personal experience. Labov points out that evaluation is the means used by the narrator "to indicate the point of the narrative, its *raison d'être*: why it was told and what the narrator is getting at" (Labov 1972). Thus evaluation can also be seen to have a significant function within a narrative.

The student interviews in our own data are similar in some ways to those collected by Chafe and Labov in that they involve the recall of personal experience. Because they involve a recounting of experience, they contain both narrative and evaluative elements.

However, the interviews we collected from the students also differed in important ways from those of Chafe and Labov. Perhaps the major difference is a difference in overall function. Briefly, the overall task of evaluation in the narrative situation is to establish that a given story is worth telling, or that a given fact or event is indeed remarkable. The "burden of proof" that his story is a good one lies with the person telling the story. The "burden of proof" in the case of our students, however, involves more than evaluating an experience to show that it is "interesting." The nature of the student's task in responding to the prompt in this assessment ultimately involves the production of another genre: persuasive argument. In particular, it involves the framing of an argument in such a way as to answer one or more of the following questions: (cf. prompt analysis above) 1) why the student chose a particular problem above other possible choices, 2) why the problem was generally relevant (i.e. why the problem was a problem), and/or 3) why the problem was personally relevant. In the students' productions, both oral and written, these "arguments" appear in the form of discussions of why certain problems are more worthy of consideration than others. The nature of the students' task required an evaluative "frame" for the purpose of argument to which reference to personal experience could be applied. Thus, although "narrative" does occur in the form of the recounting of personal experience in our data (incorporating the sort of development analyzed by Chafe and Labov), narrative elements generally served secondary illustrative functions.

In examining our data, we identified idea units with functions which are similar to the function of a narrative clause (cf. Labov 1972). These appeared in loosely coherent groups of idea units, with evaluation embedded within the idea units, or with evaluative commentary framing the identifiable group of idea units.

These groups of idea units usually appeared in the form of discussion of a series of events or observations, more or less in order of occurrence. In general, they served the function of illustrating through the recounting of actions, one or several points within a larger evaluative structure. Although these groups resemble narratives in structure, they are used to "make a point," within that larger structure. Thus while narrative embeds evaluation in the data analyzed by Labov, in our own data the reverse occurred. Narrative was embedded in evaluation. The category "Reference to Personal Experience (see above), could thus be considered a sub-category of Evaluation.

In the following example, the student evaluates the topic he is considering, and then proceeds to relate a series of events as an illustration of his point:

(25) . . . but . . . oh I know, ok, I know one that's good here,

In the process, he is filling in necessary information for the interviewer, and it is recounted as a series of temporally ordered events. The effort to use these events to illustrate a point requires that the student link his past experience to his present task. Thus a transition between two worlds occurs in idea unit 25.

- (25) . . . but . . . oh I know, ok, I know one that's good here,  
(26) . . . that I ca . . . I came here in the beginning of the year  
(27) . . . and I had a ca . . . I had my schedule,  
(28) . . . We got to pick our schedules.  
(29) . . . And so I had my schedule.  
(30) . . . And . . . the geometry class I was in was too big.  
(31) . . . So they took . . . they took me out,  
(32) . . . transferred me out of my geometry class,  
(33) and put me in with the sixth period,  
(34) . . . which means . . . which meant I had to change my English class too. (Don, Interview)

Narrative-like segments were also embedded within the evaluations of plans for composing, as can be seen in the following example:

- (15) . . . How would I go about writing it. (laughs)  
(16) . . . Probably . . . act like,  
(17) . . . um . . . write it more like a story..

- (18) . . . Like we're um . . . just rushing out of fourth period,  
 (19) . . . and . . . you've gotten all your books.  
 (20) . . . For your next class.  
 (21) . . . Run down the halls.  
 (22) . . . And to stand in a big mile long line.  
 (23) . . . I'm exaggerating a little bit,  
 (24) . . . but . . . you know . . . to get the print across.  
 (Katie)

Labov defines evaluation as "that part of a narrative which reveals the attitude of the narrator towards the narrative by emphasizing the relative importance of some narrative units as compared to others" (Labov, 1978, p. 37). He also distinguishes between degrees of embedding of the evaluation in the narrative framework, with symbolic actions being the most "internalized" type, and "direct statements of the narrator to the listener about his feelings," being the most "external." Idea unit (2<sup>o</sup>) above, ". . . I'm exaggerating a little bit," corresponds to Labov's notion of "external" evaluation.

Evaluative Strategies. Besides the overtly evaluative comments such as the one in idea unit (23) above, we found that students adopted a variety of evaluative "strategies." The following types were prevalent in our data:

1. Attribution
2. Generalized Attribution
3. Generalization
4. Comparison

These strategies appeared in both the oral and the written data. Several differences between the oral and the written data were revealed in our examination and will be discussed below. Briefly, however, they include a difference in audience, a difference in situation, and a difference between oral and written language, per se. These differences complicated our task of cross-referencing between the two sets of data. However, in terms of content, related groups of idea units were readily identifiable.

Attribution. Speakers and writers make judgments in the development of an argument. As language users, our general disposition is to categorize and to characterize according to experience. Any event has the possibility of being represented in different ways through the assignment of different roles to the people

involved. This process operates at several levels. For example, at one level, we can assign the roles of antagonist/protagonist. At another level, we can assign semantic-syntactic roles, which form contrasting pairs such as agent/patient. In some cases, assignment of roles is purposeful and serves an evaluative function.

In the process of examining our data, we found that many of the students attributed opinions and/or experience to others. In assigning opinions or experience to others while recounting events, the student provides additional support for his point of view, thus "bolstering" his argument. Generally speaking, the students' assignment of roles to individuals and groups appeared to have an evaluative function.

In our data, when students assigned opinion or the role of "experiencer" to someone other than themselves, the result was somewhat similar to what Labov has described as "embedded" evaluation. In discussing evaluation in narrative, Labov cites two examples of evaluation that are deeply embedded. One example involves a narrator addressing a second person within narrative-internal time (Labov, 1972, pp. 372). Corresponding to this case, some students in our data referenced a second person to indicate that their opinions were not exclusively their own. This device creates a consensus group to support their arguments:

(49-51) Cause my friend he always says . . . it's  
a pain. You never get what you want or anything.  
(Chip, Interview)

\*\*\*

(18) A friend and I were discussing another way of  
solving the litter problem  
(19) and we came up with this suggestion.  
(Carl, Interview)

\*\*\*

(13) A lot easier for me and my friends,  
(14) And . . . well, I guess my friends are annoyed too.  
(Don, Interview)

\*\*\*

Labov's second example shows the introduction of a third person making an evaluative comment, which has the effect of increasing dramatic force because the comment comes from a "neutral" observer. One student in our data wrote:

- (15) One other reason I chose this subject was because my brother also had problems with the open classroom.

In several instances, students provided additional support for their arguments by referencing not only their own experience, but also that related to them by relatives or friends. Thus, while Labov's examples illustrate evaluation embedded in the comments of the narrator to characters (or the comments of characters to the narrator) in narrative-internal time, our own data revealed embedded evaluation in the form of "shared" opinions and "shared" experience.

Generalized Attribution. In examining our data we also found many instances of embedded evaluation where whole idea units were used to introduce larger consensus groups. We have named this strategy "generalized attribution" to distinguish it from the strategy of attribution where consensus groups are limited to specific individuals (or groups of individuals). These strategies were signaled most clearly by mass or plural nouns (kids, students, people), plural pronouns (they and we) indefinite pronouns (everybody, everyone), and the generic pronoun (you), which generalizes the attribution to include anyone confronted with the situation being described. Generalized attribution can attribute either negative or positive characteristics to large groups sharing "consensus." In the following example, Carla places herself within a positive consensus group, closely aligned to or synonymous with that group to which she attributes an authoritative, decision-making role:

- (43) People have tried a lot of ways --  
\*\*\*

- (52) We've tried a lot of things -- and . . . nothing seems to get through their heads. (Carla, Interview)

In another example, Martin attributes to "students here" an "attitude problem" which, he claims, is at the root of all the problems he learned about in connection with planned improvements at the school.

- (13) . . . like the way they're above everything else.  
(Martin, Interview)

Students seem to establish negative attribution groups for the purpose of aligning themselves in opposition to these groups. To establish support in their opposition, they may create positive attribution groups.

- (21) It would work by student ignoring other student who don't care and telling them that they do care what happens to their school. (Martin, essay)

While the assignment of roles contributes to "contrasting" evaluation, some cases of generalized attribution achieve the reverse effect, the impression of universal agreement. Generic pronouns "you", "one", "everyone", "everybody" are abstractions, markers of the speaker/writer assuming that, within the particular conditions being described in the generic present, all people would react in the same way. The generic pronoun is often accompanied by a conditional predicate frame or some other comparative context, e. g., negative, future, question, (c.f. discussion on comparative evaluation), as in the following examples:

- (24) . . . and then . . . you sort of . . . lose friends. (Joe)  
\*\*\*  
(18-19) . . . At home you wouldn't have people/. . . you know  
. . . throwing things on the ground and---, (Carla)  
\*\*\*  
(6-7) When you are distracted from the classroom, you will  
end up in trouble/(8) and your grades will drop  
drastically.

Related to both attribution and the following discussion of generalization is the use of an inclusive, generalized attribution to "everyone" and "everybody" which functions to expand an established reference to a consensus group including the narrator: "we", "everyone" and "everybody" mark the consensus group "we" as all-inclusive, or without exception. "Everyone" and "everybody" in the following example(s) are accompanied with the generic present tense.

- (15) . . . whereas everybody . . . EVERYBODY at Central  
(16) . . . I mean everybody wants . . . to do it the way we've  
been doing it, (Alice, Interview)  
\*\*\*  
(22) . . . Because that seems to give the most problem to  
everybody. (Jay, Interview)

Another variation of this general type involves the introduction of a hypothetical "individual" representing all students within the same set of circumstances.

- (21) By changing a student's class and teacher it first makes it hard for the student to adjust to new teaching methods. (Don, essay)     \*\*\*
- (17) One would think that a student would be interested in how his school functions. (Kim, essay)

In this last example, this "hypothetical individual" became the main strategic vehicle adopted by the student in developing her argument.

Generalization. Generalization forms a third type of evaluative strategy which is characterized by a modification of a category to widen its natural boundaries. This alteration allows the speaker/writer to avoid direct statement or specification, and thus to accommodate the audience's capacity for assimilation and agreement with conditions specified by the speaker/writer. Generalization, when applied to subjects or objects, appears in the form of: dummy subjects/objects, mass nouns and pluralized sets in a generalizing (nonattributive) context.

Carla uses the dummy subject in her essay:

- (3) It is really incredible how it's so difficult to pick up a piece of paper and put it into the garbage can. (Carla, Essay)

and mass nouns in her preinterview:

- (30) . . . a lot of people think, (Carla, Interview)

Martin uses a definite article with a favored mass noun:

- (47) I think the people in the . . . in the United States,  
(48) . . . the people in the school are the most.  
(Martin, Interview)

When applied to actions, generalization transforms actions into conditions or states characterizing the subject, appearing in the form of the generic present and the progressive participial. Progressives serve to describe the action of the modified verb as constant and ongoing.

- (10) Like all these people . . . standing in front of me  
(11) ~~Representing~~ something (Joe)  
\*\*\*  
(21) All the people ~~are coming~~ here  
(22) . . . and ~~standing~~ other schools up. (Don)

Finally, "hedges" occur within the student data, serving to generally obscure the boundaries of a category by expanding ("a lot of") or constricting ("some," "many," "sort of," and "kind of") to fit the purposes of the speaker/ writer at the point of his discussion.

Obviously, more features belong to our category "generalization" as we have defined it. We have restricted our analysis and discussion to the most readily identifiable examples found in the data. Before directing our attention to the next evaluative category in our discussion, we should note here that generalizing modifiers appear in "clusters," and often serve to "qualify" evaluated material, such as in the following examples:

Joe evaluates his identification of a problem with one mass noun, one pluralized noun and a generic present tense form of "to change":

- (16-17) . . . and the way . . . people get into groups --, . . . and change.

Martin evaluates his problem by using the "hedge" "kind of":

- (12) I think the . . . students here have a . . . kind of a attitude problem. (Martin, Interview)

Jay, balancing the pros and cons of his choice of problem, uses two hedges and the generic present tense:

- (14) But then uh-- . . . you have so much noise around,  
 (15) . . . that it kind of uh . . . disturbs your uh . . . thinking.  
 (17) . . . you kinda really get that . . . cooped up feeling  
 all over. (Jay)

These instances of generalization serve the function of qualifying evaluative commentary.

Comparators. Labov's category of evaluative "comparators" has been adopted, intact, in our analysis and discussion. According to Labov, comparators include the following forms: negatives, futures, questions, or-clauses, superlatives, and comparatives. Introducing his discussion of comparatives with negative forms, Labov describes these as representing the defeat of an expectation "that something would happen". "Negative sentences draw upon a cognitive background considerably richer than the set of events which were observed. They provide a way of evaluating events by placing them against the background of other events which might have happened, but which did not occur."

Examples in our data indicate that this is indeed the case, especially when the negative is used in combination with the conditional aspect:

- (31-32) . . . you know . . . people sit here, . . . and there  
 . . . you can't sit with them. (Joe, Interview)  
 \*\*\*
- (6) Unfortunately, many students and teachers don't  
 feel as fortunate as I do. (Joe, Essay)  
 \*\*\*
- (9) . . . because um . . . litter seems to be a big  
 problem when it really shouldn't be.  
 (Carla, Interview)  
 \*\*\*
- (18) . . . At home you wouldn't have people,  
 . . . you know . . . throwing things on the ground  
 and--. (Carla, Interview)  
 \*\*\*
- (48) and kids just don't show up. (Kim, Interview)  
 \*\*\*
- (30) . . . it wasn't landscaped then,  
 (31) . . . and every time I've been at school it's never been  
 landscaped. (Bob, Interview)  
 \*\*\*

- (41) . . . My grades haven't been too well,  
 (Dor, Interview)  
 \*\*\*
- (26) They won't let us do it. (Alice, Interview)

The following is a particularly plaintive case of embedding of defeated expectation:

- (16) I have firmly believed that many teachers don't like helping the students. (Joe, Essay)

While generalization usually serves to expand a given category, comparing and contrasting through modality essentially serves to specify the conditions under which a category is presented. This is especially true when futures and conditionals are associated with evaluatively marked commentary. Fillmore's description of modality as it functions in case grammar provides us with the clearest delineation of nodal function: "modals serve to overtly mark an inference that the speaker is making . . . when a speaker shifts mood, he indicates something is likely (will), possible (may), desirable (should), obligatory (must)" (Fillmore, 1980). It follows that conditional modals, would and could, imply conditional likelihood and possibility respectively. In investigating our data, we found that the use of conditional modals in both the preinterviews and the essays was pervasive. Within the preinterviews, conditional modals were largely based in the hypothetical nature of the task in which the students were engaged (describing what they would write). Typical examples include the following:

- (7) . . . And . . . this is probably what I'd write about . . . because--
- (8) . . . well one of the things I would write about, (Carla)  
 \*\*\*
- (40) . . . And . . . that would probably be the um . . . second paragraph,
- (95) . . . And that's probably how I'd . . . close off. (Kin)

In other cases, conditionals serve to conditionalize the solution of a problem to the student's suggestion. These are often framed within a causal if-then construction:

- (11) I think if we had an enclosed cafeteria, (Carla, Essay)  
\*\*\*
- (85) . . . If . . . you know . . . we could get more . . .  
student participation.  
(86) . . . we could probably . . . solve most of these problems--  
(Carla, Essay)  
\*\*\*
- (7) because if the student had a better attitude  
toward their school  
(8) it would be a better school.  
(9) If the student gave so respect to the school  
(10) it would do more than anything else could.  
(Martin, Essay)  
\*\*\*
- (22) Then if you do this  
(23) you would improve the school more than landscaping  
more than C.I.P. (Martin, Essay)

These if-then and comparable (when-then) constructions appeared in several essays. Perhaps the most interesting cases of modal usage involved if-then framing devices qualifying the proposed institution of the student's suggestion:

- (2a) The suggestion that I would make  
(3) If I was asked  
(2b) Would be the attitudes of people at Central. (Joe, Essay)  
\*\*\*
- (14) The basic goal that I would have is trying to work  
together and get everything pulled together.  
(Joe, Essay)

Rhetorical questions also fit a discussion of evaluation and are also represented as a rather sophisticated evaluative device produced by Kim in her essay. In rhetorical questions, the answer is "planted" within the form of the question and the question is addressed directly to the audience with the assumption of agreement being a function of the form. Associated with her use of the device is the conditional frame signaled by "if", and the hypothetical student established in previous sentences:

- (26) If a student doesn't bother to take pride in his school, know how it functions, and hold interest in activities,
- (27) how do we know he will develop an interest later on?
- (28) Will he pick up litter in his office,
- (29) or will his office resemble a pigsty, reflecting on his work?
- (30) Will he register to vote for the president of the United States,
- (31) or will he choose to let other people elect the president for him? (Kim)

### Comparison of Oral and Written Units

An underlying assumption in our analysis was that the student is accessing information in the process of reading the prompt which he can then use in composing a written response. The construction of an interpretation of the prompt is thus viewed as an integral part of the composing process. In order to determine how students used their initial interpretations in their written compositions we compared segments of the oral interviews to segments of the written compositions. Both similarities and differences were found between the oral language of the interviews, and the written language of the students' compositions. Basically the similarities were found in content and the differences were found in rhetorical function.

Differences between oral and written language have received much attention in the past decade by linguists working in the field of text analysis (Ochs, 1979, Chafe, 1979, Tannen, 1978). Much of this work has focussed on the ways in which the same ideas, concepts or events, are expressed differently in written and spoken language, with contrasting descriptions of the characteristics of the language used. Tannen has noted, for example, that "...planned written discourse makes use of complex syntactic structures, formal cohesive devices, and topic sentences. In general it is more compact" (Tannen, 1978). Chafe has also commented on this characteristic of written language, in describing it as "integrated." Spoken language in contrast, is often loosely strung together with coordinating conjunctions (or with no overt marking of the relationships between propositions whatsoever) (Chafe, 1979). The result is, that ideas, descriptions of characters, and/or events which are typically expressed in several loosely joined clauses in oral language are often expressed in fewer words in written language because of the greater degree of integration one is able to achieve in writing through processes of subordination and subject deletion.

In our data, differences between oral and written units reflecting common content were both a function of the transformations which occur between oral and written styles and a function of the necessity to perform within two different situations. One situation involved the student interacting with both the prompt and an interviewer. The other situation involved writing an essay in response to the prompt which would be evaluated by the teachers at the school.

An example of the sort of compression that occurred in our data follows. The student's oral comments are his initial responses upon reading the prompt. The first idea unit makes it clear that the student is referring to the first sentence of the prompt. Later, he refers to the second sentence (idea unit 11):

- (2) . . . Ah, . . . the first thing about improvements,
- (3) Y'know like, that I like, ah . . . landscaping,
- (4) they are thinking about . . . um
- (5) . . . giving money for . . . to make the school look a little better.
- (6) And the--
- (7) . . . they've . . . they've got a big grant,
- (8) . . . state or federal or something,
- (9) . . . to improve the curriculum
- (10) . . . of the school.
- (11) . . . I know, the thing about a . . . situation or . . . problem,
- (12) I think the . . . students here have a . . . kind of an attitude problem. (Martin, Interview)

In Martin's written composition, the "same" topics are covered in his opening sentence:

- (1) The many suggestion for improvement have been made landscaping, the C.I.P. Grant, and the addited (sic) of the students. (Martin, Written)

While a total of 72 words are devoted to providing a background to the introduction of these ideas in the oral interview, only 18 are used in forming a list in the written composition. The compression that occurs in these examples is partly due to differences in the characteristics of oral and written language. However, compression also occurs because the audiences of the oral interview and the written composition are different. In the oral interview, the comments are addressed to an outsider, a "naive" interlocuter, who

doesn't know, presumably, about recurring problems, recent suggestions, or recent acquisition of grant monies at the school. Thus much of this section functions to accommodate the interviewer, i.e., to acquaint the interviewer with the facts.

The two segments also serve different purposes. The oral segment, as noted above, serves to identify for the interviewer what has been left unsaid in the first sentence of the prompt. It also signals a process of "contextualization"\* by the student (in his initial reaction to the prompt). In the written version, on the other hand, the background of information is condensed into a list identifying suggestions made. It is interesting to note that Martin's chosen problem appears within the list of established suggestions. By listing references to the first sentence of the prompt, Martin demonstrates his awareness of the "many suggestions". By framing his own subject within the list, Martin may be hoping to immediately validate his choice as one of an established set of issues that others have raised. The segment from Martin's written essay suggests that he regarded a proper response to the prompt as including an identification of the suggestions mentioned in S1 and an introduction of his own topic in relation to those suggestions. The condensation of the background detail found in the oral idea units into the succinct list in the written counterpart suggests that Martin read sentence 1 as a presentation of old information to be identified as part of the task he was to perform.

Strategies adopted by the students clearly shifted as a function of how they regarded their task and oral versions of ideas that occurred to the students in the process of reading and interpreting the prompt were not always longer than their written counterparts. Oral versions of ideas were often expanded in their written counterparts. In contrast to Martin, Katie never contextualized the first sentence with background information. She identified lunch lines as a problem at the school in her initial response, and then made the following comments in describing how she would "go about" writing her paper.

- (18) . . . Like we're um . . . just rushing out  
of fourth period,  
(19) . . . and . . . you've gotten all your books.

\*"Contextualization" refers to the process whereby a reader (or speaker) creates an imagined context for a particular utterance in comprehension.

- (20) . . . For your next class.  
 (21) . . . Run down the halls.  
 (22) . . . And . . . to stand in a big mile long line.  
 (23) . . . I'm exaggerating a little bit,  
 (24) . . . but . . . you know . . . to get the point across.  
 (25) . . . That's how I'd write it.  
 (26) . . . And then . . . standing there,  
 (27) . . . for half the . . . lunch period,  
 (28) . . . wait for you lunch.  
 (29) . . . (laughs) And then you finally get your lunch  
 go outside and eat it.  
 (30) . . . And then . . . the period'd be over while you're  
 still eating your lunch.  
 (31) . . . s . . . And then . . . you could finish your  
 lunch and then go to class--  
 (32) . . . like P.E. (laughs) . . . It's not nice to go  
 to class when you've just hurried up and eaten your  
 lunch. (Katie, Interview)

In her written version, Katie devoted many more words to covering the same material: what is originally treated in 14 idea units in the oral interview is expanded to 25 segments in the written.

- 1 Buzzzzz.  
 2 The five minute bell rang.  
 3 My fourth period teacher had finished his lecture  
 4 and we had the rest of the period for ourselves.  
 5 I scrawled out one last math problem before closing  
 my books.  
 6 "How much time, Kirsten?"  
 7 I asked  
 8 "Two minutes."  
 9 Kirsten said after looking at her watch.  
 10 I put my homework paper into my binder,  
 11 and stacked my books in my arms.  
 12 Bzzzz.  
 13 I jumped up  
 14 and raced out of the room merging into the traffic  
 of other students hurrying to lunch,  
 15 I ran to my locker,  
 16 shoved in my books  
 17 and sprinted to the cafeteria.  
 18 There I was met by a gigantic line.  
 19 My stomach growling I waited in line for  
 fifteen minutes.  
 20 With a sigh of relief I got my lunch  
 21 and went out side (sic) to eat.  
 22 Half way (sic) through my lunch the bell rang  
 signaling (sic) for the end of lunch.  
 23 This time my stomach didnt (sic) growl,  
 24 I did!  
 25 This situation is a slightly exaggerated  
 example of one of the main problems at Central High. (Katie, Essay)

The different rhetorical strategies in the oral and written segments are related to differences in the tasks being performed, (i.e., responding to an interview situation vs. composing a written text) and to differences in the audiences being addressed. In the oral version, for example, Katie chooses to give a sketchy outline of her plan, while in the written version she has to actually implement it. In the oral version, she makes use of the generic pronoun you, implying that the experience is typical, one that happens everyday to anyone. Her use of the generic you generalizes a personal frame of reference or experience, and the experience serves as an illustration for the interviewer. In the written version, different rhetorical strategies are used. The event is dramatized as a "real" event that takes place sometime in the past, the author takes the role of protagonist and the narrative proceeds from her perspective. An additional character is introduced as well as dialogue. Most of the expansion occurs in lines 1 through 12 in the written text, where Katie is introducing new elements of detail in creating a context for the episode she wishes to present. Although different rhetorical strategies are adopted, the underlying structure of the event is the same in the oral and written segments. The "same" event sequences occur between lines 13 through 18 in her written text, and between lines 18 through 22 in the interview. Comparable sequences also occur between lines 19 through 22 in her written text and lines 26 through 30 in the interview.

Katie is the single example in the subsample of eleven students of a competent writer, who is aware of the positive effects a vividly expressed personal "contextualization" can have on the point one is trying to make. She begins her composition by embarking immediately on an experience rooted in a personal perspective, complete with sounds, actions, and feelings described in sequence. Once this contextualization is completed, it testifies for her argument.

Our purpose in presenting her oral and written data is not to discuss the merits of a successfully contextualized argument, but to demonstrate how one student, without referencing the first sentence of the prompt, nor the precedent of arguments and suggestions it presupposes, proceeded to construct a successful composition. Katie creates a single imagined world with a sequence of images from her own experience. She

plans her presentation in the interview, and sticks to her plan in her composition, expanding her initial ideas to deepen and enhance her narrative. She effectively "ignores" the first sentence of the prompt, and consequently is able to argue from the strength of her own perspective. The result is that her essay gives the impression of maintaining a consistent voice with a strong argument.

Martin, in contrast (See Martin's interview above), provides an example of a student who feels it is necessary to justify his choice of a problem relative to what he knows to be a defined set of suggestions that had been made for school improvement. This is reflected in the fact that he lists those suggestions in the beginning of his essay, and compares his choice of subject matter to the other suggestions at several points in his essay:

- (9) If the student gave so (sic) respect to the school  
(10) It would do more than anything else could.

\* \* \*

- (12) By bettering their out look (sic) on this school  
It would improve it.  
(13) Improve it more than C.I.P.

\* \* \*

- (16) And what is landscaping  
(17) Just a superficial (sic) cover that would be  
destroy (sic) by student (sic) with bad  
attitude. (sic)

\* \* \*

- (22) Then if you do this  
(23) you would improve the school more than landscaping  
more than C.I.P.  
(24) These other thing (sic) could only be fully  
apprecated (sic) by student (sic) who care about  
their school. (Martin, Essay)

## DISCUSSION

The examples above illustrate the complexities we encountered in the data. Ideas which occurred to the students could be traced to segments of the written compositions, but they varied in length from one set of texts to the other, they varied in sequence, they varied in terms of the characteristics of the language used, and they varied in terms of the functions they served in relation to the task at hand.

A similar variability in the ways subjects divide the "same" objective information is recounted in Chafe's discussion of the deployment of consciousness

in narrative (1930, pp. 1-50). Although most of our data was not narrative in form, Chafe's discussion provides a partial explanation for the variability we found. Chafe notes that "centers of interest, as expressed in extended sentences, show significantly greater variation in the amount of information they contain than do focuses of consciousness, as expressed in idea units" (Chafe, 1980, p. 38). Chafe uses the term "centers of interest" to refer to linguistic units which consist of a series of focuses of consciousness and allow us to express information that would be too extensive for expression in a single idea unit. Chafe also notes that "a particular span of experience does not necessarily dictate a particular division into centers of interest." Thus the same individual may divide the "same" information in different ways at different times, either by incorporating different kinds of information into a single center of interest, or by dividing larger chunks of information into different sequences of such units.

Some of the variability we found in the ways the "same" information was expressed differently in the oral transcripts and the written compositions may thus be explained by the notion that speakers and writers make judgments in the process of verbalization, and that such judgments can vary from speaker to speaker, or from writer to writer at different times. Given that the two instances of verbalization in our data were also addressed to different audiences, occurred in very different contexts (an oral interview and a writing assessment), and were conveyed in two different mediums, it is not surprising that a great deal of variability existed. This variability prevented a strict one-to-one accounting of the "idea units" occurring across the oral and written sets of data. Instead, we independently identified larger sequences of units, reconciled discrepancies, and adopted a descriptive analysis. For most of the students, however, it was clear that the ideas that occurred to them during their initial reading and interpretation of the prompt were used in one form or another in the essays.

In discussing the results of our analysis, it is important to point out that the analysis was not based on a single reading of the prompt. Several times during the interviews, students were prompted to look back at the prompt, read it again, and to think if there were any additional things that came to mind. Even though the students had been selected by their teachers as students who would be likely to be comfortable in an interview situation, several of the students required repeated prompts before they could focus on a topic for their essay. It is entirely likely that this prompting

had an effect on the students' process as they called upon personal experience and knowledge of conditions at the school to meet the perceived demands of the task. Thus the "ideas" that were generated in the oral interview might not have been precisely the same ideas the the students would have generated had they been reading the prompt in isolation before writing their interviews.

It is also important to point out that several of the students commented that they felt they had an advantage over the other students in the assessment, because they had time to think about their composition in advance. It is difficult to calculate the effects of interviewer prompts and increased planning time on their performance in the assessment.

Clearly, the limitations in any sort of protocol analysis include the possibility that researchers interfere with the very process they are attempting to study. Increased training in the "think-aloud" procedure, so that prompting would be less necessary, would be advisable in any future study. In addition, interviewer prompts should be standardized, so that variation in individual interviewing style is minimized.

Although the conclusions that can be drawn from this analysis are somewhat limited, the data provides some insight into the ways meaning is constructed as students fit their own knowledge and experience to perceived task demands, diverging at times from the intended meaning of the authors of a prompt in the process, and transforming meaning even further as they implement their ideas in composing written text. For these students, however, the task of responding to the text of a prompt and composing a written composition were closely related. The data indicate that both reading and writing are intertwined in the composing process, that some variation in individual task constructions can be traced to ways students interpreted prompt segments, and that some variation is the result of rhetorical choices the student makes when composing a response. The data also indicate that the students in the subsample made various interpretations of prompt segments, and that these interpretations were related to the rhetorical strategies they adopted in composing a response to the prompt.

## CONCLUSION

In Part II of this report we discussed the process of prompt construction in a prompt development session involving three teachers. The original text of the

prompt was introduced by one of the teachers in the following form:

There are currently plans being formed for improvements to Central High School as a result of a State of California grant for School Improvement. There are many committees working on plans for school improvement. Identify three improvements you would like to see made at Central, telling why you feel each improvement needs to be made and what you would like done.

This version was immediately changed because it was found to be too structured. The teacher commented:

. . . I didn't wanna go into the whole . . . grant . . . that we got . . . I think most of the kids know about it . . . now . . . I've structured it too much . . . so I'm gonna change it as I read . . . . \*

The original prompt author in a subsequent communication to researchers offered further explanations for why the original wording of the prompt was changed:

It struck me, as I began to read it as being unnecessarily formal and wordy (Did kids really need to know all the background in order to write? If they did, could it be put in simpler form?)

During the session, specific reference to available grant funds and other references to the precise context the authors of the prompt had in mind were ultimately dropped in favor of what they regarded as a more general statement, intended to reassure students that there was no right answer to the prompt directives contained in sentence 2. The authors of the prompt did not feel that the "specific" context they originally had in mind would present a problem to students who weren't privy to recent discussions within committees reviewing proposals for use of recently acquired grant funds. They appeared satisfied that they had sufficiently generalized their initial context to allow for a wide range of possible responses from the students.

It is important to emphasize at this point that

the authors of the prompt began with the description of a very specific context, and were attempting to move away from specification of that context toward a simpler subject. The result of their efforts was sentence 1, containing a residual core of information removed from the specific facts and circumstances surrounding it.

Many different suggestions for improvement of Central High School have recently been made.

Teachers commonly construct prompts to include initial "contextualizing" or backgrounding material. In many cases, as was the case here, teachers view these contextualizing statements as secondary to the task indicated in the remainder of the prompt. It is not surprising therefore, that the additional group of holistic scorers, like the developers of the prompt, did not predict the consequences of S1. It would be erroneous to assume, however, that all students would view these contextualizing statements as secondary given the conditions of an assessment situation. Sentence 1 is recognizable as a condensed version of shared information. Thus, it is understandable why some students would construe S1 as their cue to reconstruct the referenced context and proceed to write from the perspectives existing within that context. Students who were preoccupied with the reconstruction of the "given" context for S1 found it difficult to adopt a personal perspective and tended to define their "voice" for the purpose of writing this essay in terms of opinions shared by others.

The analysis described in Part II of this report indicates that the target sample of students made different interpretations of segments of the prompt. These differences have the potential for affecting the evaluation of student performance if they lead the students to adopt strategies for writing which do not match the expectations of the evaluators. In the case of the students in this sample, ideas generated in the course of reading the prompt appeared in some form in the written essays. In addition, the strategies for writing adopted by several of the students in the subsample involved in large measure the sort of "evaluation" strategies described above. In this data, "evaluation" did not have the positive effect that it is described as having in narrative. If anything, the reverse was true, for the students often failed to present an "objective" case, and relied instead on the rhetorical device of attributing opinion to a second person ("my friend says") or to a generalized group

("kids say"). One could argue, that these strategies depended in part on the prompt, since "describing a problem" is a potentially controversial subject, at least from the student's point of view. The prompt requires students to "criticize" their school, which may lead some students to select "safe" subjects and in general fail to adopt a strong personal voice.

The extensive evaluative commentary in the student essays reflects their lack of confidence in the relevance and appropriateness of their own opinions. In our analysis of the prompt, we emphasized that S1 contained the basis for misconceptions on the part of students of what the task required. We are describing these interpretations as misconceived only insofar as they did not meet the expectations of the developers for a response to the prompt. The students' interpretations of the prompt should not be considered misinterpretations. Rather, they should be viewed as variations of possible interpretations that are inherent in the prompt. Many of the students appeared to feel that their discussions should reflect previous discussions, and several of the students felt it was necessary to justify their choice and/or give reasons why they felt their choice was "generally" relevant. The following examples illustrate how one student, in an effort to account for all possible interpretations of the phrase "giving reasons for your choice," tried to accomplish two of the evaluative tasks implied in the phrase "giving reasons for your choice."

The reason I *think*\* that open classrooms are a problem in *Central* is because when the classrooms are open the way they are there is a lot of disturbances around you. (Jay, Essay)

The *reason I chose*\* this problem is because when I first started high school my grades dropped really fast. (Jay, Essay)

In the first example, the phrase "the reason I think" is followed by a conditional or logical causal relation signaled by "when," "there is," which is generalized by the use of the generic pronoun "you." In the second example, the phrase "the reason I chose" is followed by a recounting of personal experience. The former accounts for a generalized relevance of the problem chosen, and the latter expresses the personal

---

\*Italics ours.

relevance of the chosen problem.

In his discussion of the narration of personal experience, Labov views evaluation in favorable terms, and makes the point that greater complexity of evaluative devices is the sign of a developed "raconteur." While evaluation has its natural function within the narrative context, we would like to suggest that evaluative devices, when used excessively and indiscriminantly in a non-narrative expository context, can indicate that the speaker or writer is compensating for an inability to express his/her argument in a direct manner.

Our study suggests that context plays an important role in both the reading and the writing process. Clearly, a context for the first sentence (or the lack of it) influenced the students' interpretations of the prompt and ultimately the way they formulated their responses in the written essay. However, the issue is not simply one of "knowledge" or the lack of it. Rather, the issue is whether or not a context should be provided, and when. If a prompt is intended to be open, we should avoid obscured contexts which may lead some readers (such as many of the students in this study) to attempt a reconstruction of the authors' intended frame of reference. If a context of prior knowledge is considered to be important, as of course it is in many assessment situations, then prompts should clearly specify what that context is.

When students have only one chance to write, and are judged on this single performance, it is especially important that students have optimal conditions for their performance, and that there be no mismatches of either topic or task. The findings reported in Part I of this report indicate that students and teachers disagreed on many elements of task interpretation. In addition, unstated expectations were found to be correlated with score. The analysis reported in Part II indicates that various interpretations were made of prompt segments, and that these interpretations were related to rhetorical strategies adopted by the students in composing a response. The findings have serious implications for current practices in assessment, since mismatches of this kind can result in distortions of the ways individual students are judged. The first step for both the student and the evaluator in a writing assessment is the reading and interpretation of a prompt. Thus, the features of a prompt which influence interpretation are of critical importance for both the reliability and the validity of our assessment measures.

## References

- Brandt, W.J., The rhetoric of argumentation, Bobbs-Merrill Company, Inc., Indianapolis, N.Y. (1970).
- Bransford, J.D. and M.K. Johnson, "Consideration of some problems of comprehension," in Visual information processing, ed. W. Chase, Wiley, New York (1973).
- Chafe, W., "The deployment of consciousness," in The pear stories, vol. III Advances in discourse process, ed. R.O. Freedle, Ablex Publishing Corporation, Norwood, New Jersey (1980).
- Clark, H.H. and S.E. Haviland, "Comprehension and the given-new contract," in Discourse production and comprehension, ed. R.O. Freedle, pp. 1-40, Ablex Publishing Corporation, Norwood, New Jersey (1977).
- Fillmore, C.J., "Verbs of judging: An exercise in semantic description," in Studies in linguistic structure, ed. D. Terrence, pp. 273-890, Holt, Rhinehart and Winston, Langendoen, New York (1971).
- Fillmore, C.J., "The case for case," in Universals in linguistic theory, ed. R. Harms, pp. 1-90, Holt, Rhinehart and Winston, Langendoen, New York (1971).
- Fillmore, C.J., Innocence: A second idealization for linguistics, University of California, Berkeley, Berkeley, California (1971). (Manuscript)
- Fillmore, C.J., Ideal readers and real readers, Paper presented to the Georgetown University Round Table, Washington, D.C. (December, 1981).
- Fillmore, C.J. and P. Kay, A text semantic approach to testing reading, N.I.E. Progress Report, University of California, Berkeley (1981).
- Goldstein, E.O., The difference between cohesion and coherence, Cincinnati (1980). Paper presented at the First Annual Convention of the National Council of Teachers of English.
- Greenberg, K.L., Some relationships between writing tasks and students' writing performance, Boston (1981). Paper presented at the 70th Annual Convention of the National Convention of Teachers of English.
- Halliday, M.A.K. and R. Hasan, Cohesion in English, Longman, London (1976).
- Haviland, S.E. and H.H. Clark, "What's New? Acquiring new information as a process in comprehension," Journal of Verbal Learning and Verbal Behavior, vol. 13, pp. 512-521 (1974).

- Hayes-Roth, B. and P.W. Thorndyke, "Integration of knowledge from text," Journal of Verbal Learning and Verbal Behavior, vol. 18, pp. 91-108 (1979).
- Hoetker, J., Effects of essay topics on student writing: A review of the literature, Florida State University, Unpublished paper, Tallahassee (1981).
- Kintsch, W.E., W.J. Kozminsky, G. McKoon, and J.M. Keenan, "Comprehension and recall of text as a function of content variables," Journal of Verbal Learning and Verbal Behavior, vol. 14, pp. 196-214 (1975).
- Kinzer, C., Topic-based expectations of participants in the writing event, Paper presented at the Annual Meeting of the American Educational Research Association, New York (1982).
- Labov, W. and Waletzky, "Narrative analysis: Oral versions of personal experience," in Essays on the verbal and visual arts, ed. June Helm, American Ethnological Society, Distributed by the University of Washington Press, Seattle and London (1966). Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society.
- Labov, W., Language in the inner city: Studies in the black English vernacular, University of Pennsylvania Press, Philadelphia (1972).
- Leech, G., Semantics, Penguin Books, Harmondsworth, Middlesex, England (1974).
- Leech, G. and J. Svartvik, A communication grammar of English, Longman, London (1975).
- Mandler, J.M. and W.S. Johnson, "Remembrance of things parsed: story structure and recall," Cognitive Psychology, vol. 9, pp. 11-151 (1977).
- Matsunashi, A., Producing written discourse: A theory-based description of the temporal characteristics of three discourse tasks from four competent grade 12 writers, University of Illinois at Chicago Circle, Chicago, Ill. (1981). (Manuscript)
- Mehan, H., "What time is it Denise? Asking known information questions in classroom discourse," Theory Into Practice, vol. 18, pp. 285-294 (October 1979). (Periodical)
- Mehan, H., Learning lessons: Social organization in the classroom, Harvard University Press, Cambridge, Massachusetts (1979).
- Moe, A.J., "Cohesion, coherence, and the comprehension of text," Journal of Reading, vol. 22, pp. 16-20 (1979).

- Mullis, I.V.S., Highlights and trends in writing achievement, 1969-79: Results from the National Assessment of Educational Progress, Los Angeles (1981). Paper presented at the Annual Meeting of the American Educational Research Association.
- Myers, M., A procedure for writing assessment and holistic scoring, ERIC Clearinghouse on Reading and Communication Skills and the National Council of Teachers of English, Urbana, Illinois (1980).
- Pierce., Collected works, II, Bk. III, Ch. II, Pt. III, Harvard University Press, Cambridge, Massachusetts, 8 Vols. (1931-1955).
- Rumelhart, D.E., "Understand and summarizing brief stories," in Basic processes in reading, perception and comprehension, ed. S. Samuels, Lawrence Erlbaum Associates, Hillsdale, N.J..
- Ruth, L., Questions and answers: A model of relations, forms, and functions, Unpublished doctoral dissertation, University of California, Berkeley (1980).
- Ruth, L., "The state of current research and the goals and procedures of the study," Part I of Technical Report 1, N.I.E. Writing Assessment Project, Bay Area Writing Project University of California, Berkeley, California (1982).
- Sandmann, M., Subject and predicate: A contribution to the theory of Syntax, University Press, Edinburgh (1954).
- Searle, J.R., "A classification of illocutionary acts," Language In Society, vol. 5, pp. 1-23 (1976).
- Walker, C.H. and B.J.F. Meyer, "Integrating information from text: An evaluation of current theories," Review of Educational Research, vol. 50, pp. 421-437 (Fall, 1980).
- Yekovic, F.R. and C.H. Walker, "Identifying and using referents in sentence comprehension," Journal of Verbal Learning and Verbal Behavior, vol. 17, pp. 265-277 (1978).

Appendix A  
Student Questionnaire

Name \_\_\_\_\_

Teacher \_\_\_\_\_

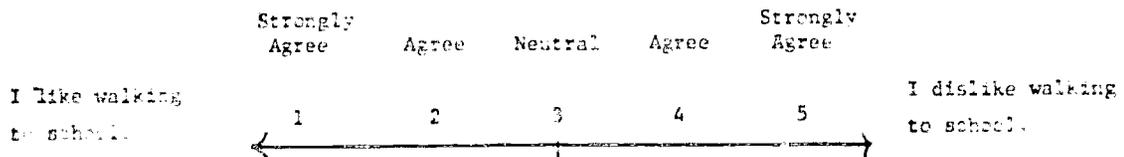
Period \_\_\_\_\_

You recently had a short writing assignment on the following topic:

Many different suggestions for improvement of Central High School have recently been made. Describe one problem or situation at Central which you feel needs correction or improvement, giving reasons for your choice and suggesting one or more solutions.

This is a questionnaire to find out how you feel about writing in general and, specifically, what you think about the topic you had to write on.

Please circle the number of the choice which best reflects your opinion about the statements in the questionnaire. Each question has two extremes, which have been placed on either side of a scale. For Example:



If you strongly agree with the statement on the left, then you would circle number 1.

If you strongly agree with the statement on the right, then you would circle number 5.

If you do not agree with one of the statements more than with the other, then you would circle the number for the neutral category -- number 3. Be sure to read the statements on both sides of the scale before you circle your choice.

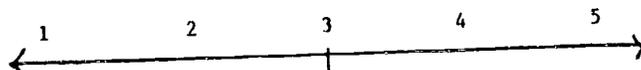
Please ask the teacher now if you do not understand what you are to do.

When you have finished, check to make sure that you have answered each question.

Be sure your name is at the top of this page.

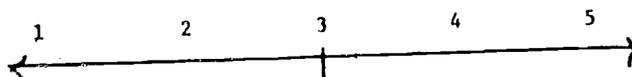
Strongly Agree      Agree      Neutral      Agree      Strongly Agree

1. I know of many (more than 3) problems which affect the entire school at Central.



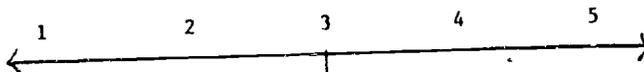
I know of no problems which affect the entire school at Central.

2. I often think about improving the school.



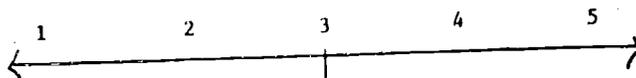
I never think about improving the school.

3. When I think of problems in the school, I think of things which affect many or all of the students.



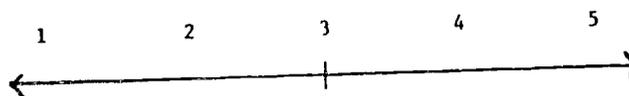
When I think of problems in the school, I think of things which affect only myself.

4. In general, Central seems to have few problems.



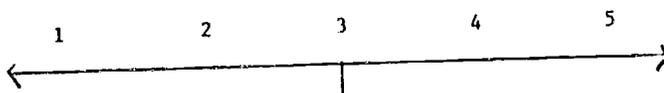
In general, Central seems to have many problems.

5. I like to suggest ways to help the school become a better place to be.



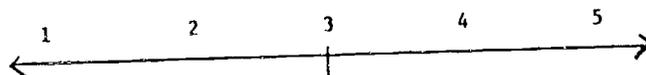
I do not like to suggest ways to help the school become a better place to be.

6. I think this topic allowed me to voice an opinion about an important issue.



I do not think this topic allowed me to voice an opinion about an important issue.

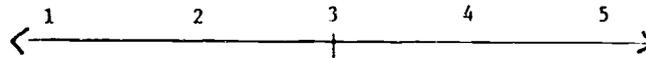
7. I am glad I had the chance to tell someone about the subject I wrote on.



I am not glad I had the chance to tell someone about the subject I wrote on.

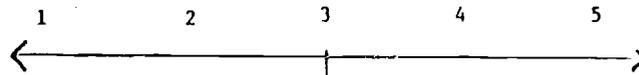
Strongly Agree      Agree      Neutral      Agree      Strongly Agree

8. The topic was a satisfying one to write on.



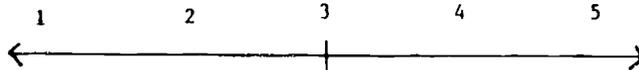
The topic was not a satisfying one to write on.

9. The topic was more pleasing to write about than most others given in school.



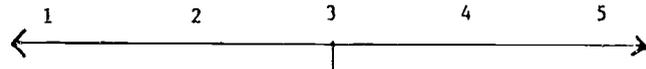
The topic was less pleasing to write about than most others given in school.

10. The topic was easy to read and understand.



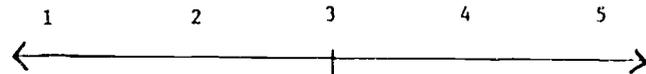
The topic was difficult to read and understand.

11. The topic was easy to write about.



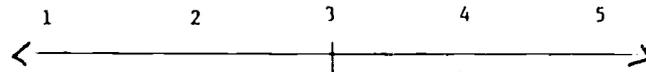
The topic was difficult to write about.

12. I had a hard time choosing something to write about.



I did not have a hard time choosing something to write about.

13. I think I wrote better for this assignment than I usually do.



I think I wrote more poorly for this assignment than I usually do.

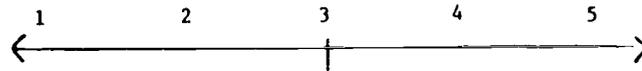
480



The following items refer specifically to the topic you wrote on yesterday. You may go back to the first page and look at the topic:

Strongly Agree      Agree      Neutral      Agree      Strongly Agree

1. I wrote on only one problem or situation in responding to the topic.

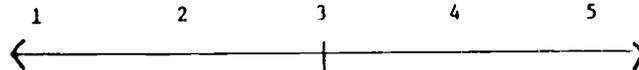


I wrote on more than one problem or situation in responding to the topic.

When Writing on This Topic:

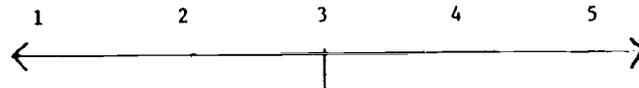
Strongly Agree      Agree      Neutral      Agree      Strongly Agree

2. It is important to describe only one problem.



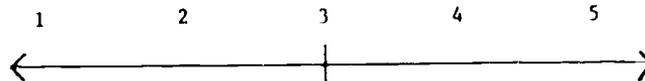
It is not important to describe only one problem.

3. It is important to explain why the subject chosen is a problem at Central.



It is not important to explain why the subject chosen is a problem at Central.

4. It is important to suggest a solution to the problem or situation.



It is not important to suggest a solution to the problem or situation.

5. It is important to suggest more than one solution to the problem or situation.

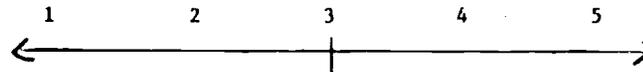


It is not important to suggest more than one solution to the problem or situation.

When Writing on This Topic:

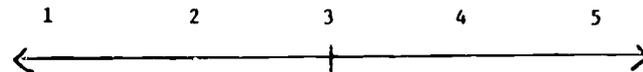
Strongly Agree      Agree      Neutral      Agree      Strongly Agree

6. It is important to point out that many suggestions for improving Central High have been made.



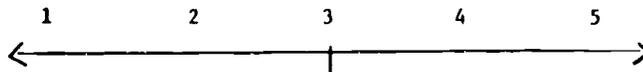
It is not important to point out that many suggestions for improving Central High have been made.

7. It is important to relate the problem or situation to everyone in the school.



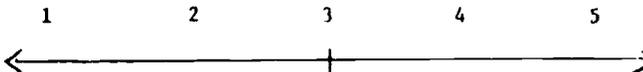
It is not important to relate the problem or situation to everyone in the school.

8. It is important to show how the problem affects the person doing the writing.



It is not important to show how the problem affects the person doing the writing.

9. It is important to explain how the solution/correction of the subject chosen will improve Central High.



It is not important to explain how the solution/correction of the subject chosen will improve Central High.

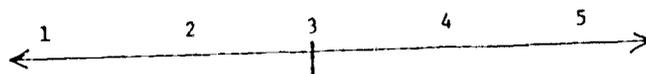
Appendix B  
Teacher Questionnaire

483

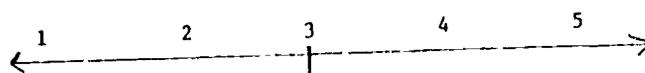
## APPENDIX B

Strongly Agree      Agree      Neutral      Agree      Strongly Agree

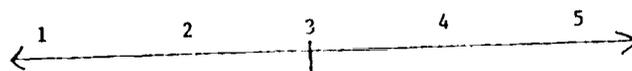
1. The topic would be more pleasing for students to write about than most others given in school.



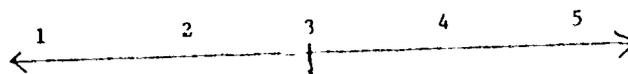
2. The topic would be easy for students to read and understand.



3. The topic would be easy for students to write about.



4. I think the topic would give a good indication of student's writing ability



The topic would be less pleasing for students to write about than most others given in school.

The topic would be difficult for students to read and understand.

The topic would be difficult for students to write about.

I think the topic would not give a good indication of student's writing ability.

The following items refer specifically to the topic. You may go back to the first page and look at the topic:

When Writing on This Topic:

- |   | Strongly<br>Agree | Agree | Neutral | Agree | Strongly<br>Agree |  |
|---|-------------------|-------|---------|-------|-------------------|--|
| 1. It is important to describe only one problem.                                  | 1                 | 2     | 3       | 4     | 5                 | It is not important to describe only one problem.                                  |
| 2. It is important to explain why the subject chosen is a problem at Central.     | 1                 | 2     | 3       | 4     | 5                 | It is not important to explain why the subject chosen is a problem at Central.     |
| 3. It is important to suggest a solution to the problem or situation.             | 1                 | 2     | 3       | 4     | 5                 | It is not important to suggest a solution to the problem or situation.             |
| 4. It is important to suggest more than one solution to the problem or situation. | 1                 | 2     | 3       | 4     | 5                 | It is not important to suggest more than one solution to the problem or situation. |

487

## When Writing on This Topic:

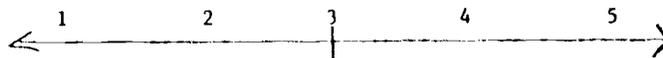
Strongly Agree      Agree      Neutral      Agree      Strongly Agree

6. It is important to point out that many suggestions for improving Central High have been made.



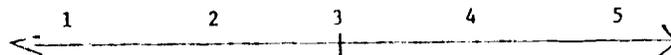
It is not important to point out that many suggestions for improving Central High have been made.

7. It is important to relate the problem or situation to everyone in the school.



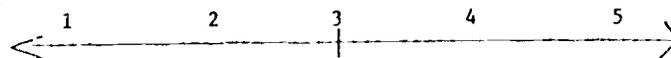
It is not important to relate the problem or situation to everyone in the school.

8. It is important to show how the problem affects the person doing the writing.



It is not important to show how the problem affects the person doing the writing.

9. It is important to explain how the solution/correction of the subject chosen will improve Central High.



It is not important to explain how the situation/correction of the subject chosen will improve Central High.

## ACKNOWLEDGEMENTS

This report is the result of the cooperation of many persons--the teachers who participated in the prompt development session and scored the essays, the students at the high school who consented to be interviewed, and the research and support staff of the Bay Area Writing Project who spent many hours transcribing the data. We wish to acknowledge our appreciation. A special note of thanks is extended to Ms. Marta Guzmán for her many hours at the word processor and to Dr. Leo Ruth for his help in preparing the final draft. We are also especially grateful for the advice given by Dr. Wallace Chafe and Ms. Jane Danielewicz on our analysis of the students' protocols and for the advice of Dr. Charles Fillmore on our analysis of the writing prompt.

187

Chapter 5

HOLISTIC ASSESSMENT AS LONGITUDINAL  
MEASURE OF STUDENT GROWTH:  
INTERPRETING STUDENT TASK CONSTRUCTIONS

Part I

UNEXPECTED DIRECTIONS OF CHANGE  
IN STUDENT WRITING PERFORMANCE

Catharine Keech

Part II

BEYOND HOLISTIC SCORING: RHETORICAL FLAWS  
THAT SIGNAL ADVANCE IN DEVELOPING WRITERS

Elissa Warantz and Catharine Keech

## UNEXPECTED DIRECTIONS OF CHANGE IN STUDENT WRITING PERFORMANCE\*

Catharine Lucas Keech

Introduction

A striking problem which recurs repeatedly in all direct measures of writing performance is the instability of student performance from one occasion to another. Users of essay tests are troubled by the frequency with which individuals and even whole groups may fail to produce consistent improvement on these measures over time--in some cases, even after concentrated writing instruction. This phenomenon has proved baffling and frustrating to teachers, researchers, and program evaluators. As children learn to write better, their scores on essay tests should simply go up. If the scores don't improve, we may be forced to conclude that students have not improved their writing abilities.

Of course, one must examine the test measures used: were the two test occasions truly parallel in what they required students to do, and were the tests scored in the same manner? Testers have reported improving performance stability substantially by improving inter-rater agreement, phrasing prompts carefully to make demands more parallel, taking longer samples to improve discrimination of the measure, and taking more than one sample at each test sitting. (Diederich, 1974; Godschalk, et al., 1976; Breland and Gaynor, 1979.) Some unexpected drops in score can be explained, in addition, by those uncontrolled conditions traditionally cited: lack of motivation, having a "bad day," failure to find good material to write about, or external distractions as when a hated substitute teacher administers the posttest on a day of campus riots in the middle of spring fever, in which case scores may drop for a whole class.

All such factors, however, are as likely to affect pretest as posttest scores: although program evaluators may examine test conditions more assiduously

*\*This paper presented at AERA, New York, 1982, provides the theoretical background for the research reported in Part II of this chapter.*

for differences which might account for disappointing gains in posttest scores, it is equally possible that any impressive signs of improvement from pre to posttest may be due to factors which artificially depressed pre test performance, creating an inflated picture of growth. In spite of these uncertainties, program evaluators in California and other states have not been driven to abandon holistic scoring of writing samples in favor of the more reliable but less valid multiple-choice test format. (Spandel and Stiggins, 1980.) But adhering to the practice of using actual writing samples in evaluating student growth in composing skills requires not only that evaluators control test conditions more rigorously but also that researchers begin to look more closely at what changes when students write on different occasions. The phenomenon of unexpected drops in scores or failure to improve in statistically significant increments on direct measures of writing is too widespread among both individuals and groups to be adequately explained either by random factors or by poor instruction--especially since, over the long run, improvement does finally seem to occur, whatever the ups and downs along the way.

Two aspects of the development of writing ability have so far, I believe, been inadequately taken into account by researchers and evaluators confronting drops in performance scores on the part of students who should perform better: (1) the general phenomenon of the U-shaped learning curve, or the existence in all complex learning of plateaus or even apparent regression as learners move from one level of competency to the next (Piaget, 1977; Goodman, 1979; Miller, 1980; Bever, 1982); and (2) the special open-ended or ill-defined nature of writing tasks which allow (even require) the writer to set for himself to some extent the difficulty level of what he attempts to do. (Bereiter and Scardamalia, 1978.)

In this paper I would like briefly to state the case against current attempts to measure growth using methods developed to measure absolute proficiency levels--measures which may be misleadingly affected by developmental unevennesses in performance. Then I will examine in some detail the problem of task definition for all longitudinal research in writing, explaining how a refined discourse typology might help us identify and describe more precisely what students do differently from one writing occasion to the next. Finally, I explore briefly new ways of asking why and how students change the difficulty level of writing tasks intended by the tester to be parallel in demands.

#### Developmental Unevenness and Current Measures

It would be odd if learning to write were somehow exempt from the roller-coaster character of learning documented by cognitive psychologists (cited above). In fact, many writing teachers have recognized some manifestations of the non-linear nature of improvement in developing writers: for instance at the sentence level, the appearance of fragments that so often accompanies the onset of subordinate clauses or noun appositives; or the run-together sentences that appear when students start developing complex ideas using sentences that stand in close opposition or apposition to one another, which they want to join with an introductory adverb rather than a subordinate conjunction. Similarly, elaborately expanded subjects, or the use of abstract nouns as subjects, frequently create predication errors or verb-subject agreement errors the student would have no difficulty avoiding in his earlier, simpler, syntax. Analytic error counts and even holistic responses may determine that these writers-in-transition are less able performers than before, when actually each of these flaws can be seen in context as signalling movement toward a higher level of competency.

Developmental evaluation is further complicated when the complexity of the cognitive task being attempted causes a writer to shift attention away from

rhetorical niceties she may have mastered but which may never become fully automatic, and so breakdown when not attended to. Freedman and Pringle (1980) point out in a study of college writers that conventional rhetorical evaluations of writing are based on factors which do not improve, and may actually suffer, when a writer is attempting to solve a communication problem that lies at the outer limits of her cognitive processing capacity. They conclude:

"In evaluating their students' writing, teachers seem to (lack) a sense of the complex nature of...development, in which growth in one dimension may entail momentary awkwardness in another...When teachers ignore the cognitive aspects of the (writing) process and focus only on the rhetorical features of the completed product,...they are not in a position to anticipate the kind of breakdown that occurs when the intellectual task is made more complex. Further,...they may actually impede intellectual growth...(The) short-run effects of such evaluation are probably to encourage students to operate on the safe levels they have already mastered..."

(p. 323)

Thus, students may write more awkwardly, less correctly, less fluently, either because they are acquiring new forms at the word or sentence level or because they are struggling to express cognitive functions for which their current language forms are as yet inadequate. A third developmental struggle has received even less investigation than these two: how the student defines the rhetorical task at the text or discourse level. In whole-text planning, what does the student set out to do, and how does he <sup>intend to</sup> do it? How does he conceive of the artifact he is about to create? What kind of text does he think in terms of, as he pursues his ideas and formulates his sentences? To what extent does he conceive of the whole text at all?

Recent research in discourse grammars and text design (Stein and Glenn, 1979; Meyer, 1975, and others) and modern discourse theorists (including Bain, 1890; Kinneavy, 1979; Britton, 1975; and others: see D'Angelo, 1976) have reestablished the importance of distinctions among genre or types of writing which the learning writer must begin to differentiate, mastering a variety of composing

strategies to accomplish different rhetorical purposes, such as reporting, explaining, persuading. Deciding what he wants to do (tell a story, explore his feelings, change someone's course of action, etc.) and selecting the appropriate composing strategies to accomplish these aims are two important parts of representing or constructing the writing task--two processes particularly subject to the writer's stage of development, which may interact in confusing ways with the writing assignment. It is in the areas of text design, rhetorical purpose and strategy, that the ill-defined nature of the writing task comes into play, creating a particular set of problems for measuring development in writing ability.

#### The Ill-Defined Writing Task

It is crucial in evaluation research to distinguish between the given task or actual text of the writing assignment--with its particular set of constraints and options, expressed or implied, and with its virtually infinite set of possible realizations--and on the other hand the constructed task, the set of constraints actually honored and the options actually chosen by the student, as seen in the text the student produces. Test makers may believe they have carefully constrained a particular task to elicit a particular kind of response, but as Murphy, Carroll and Kinzer (1982) have demonstrated, students are capable of creating totally unexpected task versions or notions of what they are "supposed to do." (See also Keech and McNelly, 1982.)

Unlike other performance areas --for instance, music, where a learner is given a more difficult piece to play, or mathematics, where both teacher and student are aware when a more difficult problem is offered, or sports competitions where more advanced dives earn more points-- composition has no such well-defined, easily ranked tasks or gradations of performance levels that can be controlled by the teacher or evaluator. It is possible

to give students the same assignment on two different occasions and have them construct such different composing tasks that their two responses are hardly comparable. The student might appear an expert on the first writing, while performing like a novice as he attempts on the second writing a far more complex problem. Sufficiently open-ended assignments, such as "Write about a favorite object," may be offered again and again over all the years of a writer's career, from first laborious printings to his adult years of proficient composing; the actual task difficulty will increase in proportion to his writing ability. The very ability to define a task more richly and complexly may be one of the most important writing skills, developing slowly along with syntax and sense of paragraphing.

Bereiter and Scardamalia (1979), recognizing that writing tasks are inherently ill-defined, are unconcerned when changes in young writers are not reflected in conventional scores on their writing:

"...(Our) most successful experiments so far in affecting children's composing processes have not led to discernible overall improvements, as judged by impressionistic ratings...This would be discouraging if the purpose were to get children to do a better job of pursuing the same goals as before. When, however, the purpose is to get children to tackle problems they have not tackled before, such impressionistic results are immaterial. It is what they are doing differently that counts, not how well they are doing it compared to how well they previously did something of a different sort." (pp. 83-84, ms.)

Evaluators trying to document growth or development in writers confront a choice: either narrow task constraints sufficiently in the writing assignment to keep writers doing the same kind of thing on trials 1 and 2, or begin to find out how to describe what children do differently from one occasion to the next, how to interpret that difference, and finally how to measure the development it implies.

So far, evaluators have quite properly concentrated on trying to make task demands similar for repeated writing samples used to measure improvement over time. **It seems obvious**

that changes in performance abilities may be obscured if a student is encouraged to write a personal experience reminiscence on one occasion and an argumentative essay on another: if nothing else, it is difficult to compare these responses quantitatively, scoring them under the same rubric. At least the very rough distinction between narration and exposition is typically made, with many program evaluators eschewing the earlier practice, associated with the Bay Area Writing Project model of holistic scoring, of allowing students to respond to a writing stimuli using almost any form of writing, sometimes including poetry.

Two dangers are inherent in attempts to narrow task constraints to reduce variation in task construal for purposes of improving comparison of samples, however. The first is simply that the more text testers add to the writing assignment, the less guarantee they have that students will read and correctly interpret all of the guidelines--in the extreme cases, students may either ignore a lengthy set of instructions, or may become so embroiled in working out exactly what the tester wants that they are distracted from their central task of trying to generate meaningful, coherent text. The second danger appears only when test instructions are so clear and so good that they actually succeed in narrowing the task and making student responses easy to compare and score because the student is being asked to do so little. Writing tests can be made to be little more than direction following, requiring students to show proficiency in the use of certain written conventions, but failing altogether to test the student's

ability to actually "author" a text, as Moffett (1979) identifies the highest composing skills. Too narrowly specified tasks maximize control over what is attempted but may invalidate the writing test, since the test no longer reveals how the student frames the rhetorical problem, sets goals, chooses strategies, juggles constraints, revises inner speech-- in short, really composes.

Ironically, it may be by looking at samples from assessments in which students were given maximum freedom to choose their purposes and strategies in response to a topic idea that we are best able to discover what kinds of writing, what discourse options are most likely to be elicited by certain topics, regardless of instructions to the writer. The early BAWP assessment model, as represented by the six-year hodge-dodge of topic types used at Sir Frances Drake High School (see Table 1), while generating scores that soar and dip in unpredictable ways from year to year, provides a rich data source for discovering patterns of task construal in response to different kinds of topics as well as in relation to grade level or experience. (See Table 1, p. 503.)

#### The Drake Longitudinal Sample

The Drake teachers who composed the test questions shown in Table 1 deliberately made them as open as possible for several reasons. They were interested in how students might construe the task: they wanted to know whether their instructional program had succeeded in teaching students the large range of strategies that might be used effectively to solve a particular rhetorical problem. Since they were attempting to teach many forms of writing, they did not want a restriction on the test to suggest that only one of these kinds was valid or important.

Further, they were unconcerned, in this pioneering effort, to make tasks similar from year to year. Initially their goal was simply to create a stimuli that would allow students to begin writing quickly, with maximum enthusiasm and ample material for composing; hence, the emphasis on personal feelings. Later they wanted to see how well students wrote evaluative or argumentative pieces (1976 and 1977.) In the last year (1978) the topic provides a complex blend of possibilities, allowing students to focus on either an experience, a person or thing, or an idea--a change in themselves.

The teachers were well aware that the 1973 topic, "Write about an event . . . ." was likely to evoke imaginative narratives, while the 1976 topic, "Name one invention we would be better off without..," would invite argument or exposition. They were neither surprised nor displeased, however, when, in 1973, many students wrote expository, choosing to identify and comment on an event they wanted to witness, rather than to relate it as if they had witnessed it; or when, in 1976, one student demonstrated the evils of television with an expert short story, while another wrote a letter to Henry Ford in heaven, berating him for having contributed to the death of the writer's parents in a car crash.

Three longitudinal samples were constructed from the Drake data base, as shown in Figure 1, each containing ninth through twelfth grade samples from 30 students. When the four papers of one student are gathered in a case study folio, it is immediately clear that students have in fact done different things on different rounds of assessment, causing their holistic scores to change in surprising ways.<sup>1</sup> Not only does the change in topic from year to year invite different kinds of writing, but students appear to set up different sorts of tasks that are not determined by topic differences,

\* Page 505.

but are allowed by the writing instructions. It is precisely the latitude allowed students that makes this sample of papers so useful in giving us an idea of the repertoire of text designs or discourse schemas high school writers have available for solving a number of different writing problems. In one sense, the Drake sample represents the worst that could happen in an assessment which is attempting to measure improvement in writing from one occasion to the next. By representing such an extreme, the sample allows us to investigate the primary problem inherent in all longitudinal writing research to a greater or lesser extent: our inability so far to account for task complexity, or to compare fairly the difficulty level of what a student attempts on different occasions.<sup>2</sup>

A system of task descriptions is needed to determine whether or not two compositions represent the same or different types of tasks: which purposes does the student appear to be aiming for? Which strategies does he use? Only after researchers are able to describe and classify the apparent underlying discourse schemes students use will they be able to distinguish between a student's mastery of old tasks and his novice attempts to accomplish new tasks. A refined discourse typology appears essential to the development and testing of hypotheses about the relative difficulty levels of what is attempted, ultimately allowing predictions about which kinds of tasks are learned first, which may follow, and what various spontaneous task construals may indicate about a given student's particular level of development.

#### A Discourse Typology for Developing Writers' Texts

Available discourse classifying systems proved inadequate for describing the student texts in the sample, possibly because most of them represent

what might be called ideal text types, the products of accomplished adult writers rather than of students who may be only approximating these discourse forms. ¶ Further, existing typologies offer three to four global categories which perforce must obscure almost as many differences between texts as they are able to identify. Just as Kinneavy suggests a distinction between the aims of discourse and the modes of discourse,<sup>3</sup> so I found it critical to separate discourse function from rhetorical strategy. To classify texts according to function is to ask: How does the text function for the reader? How might the reader characterize the writer's underlying purposes, on the basis of the whole effect of the piece? To classify strategy, the reader shifts attention from the whole to the parts: What language structures does the writer use to accomplish her purposes? The final classification of the text then is in terms of both function and strategy: what was done and how was it done?

Figure 2\* shows the three discourse functions found in the sample, together with their parallel strategies. Although each strategy appears to "belong to" a particular function, the strategies are separable from the functions, not merely by analysis but also in practice. "What to do?" and "How to do it?" are genuinely discreet questions which a writer may choose to consider separately during composing. On the other hand, some traditional discourse schemes strongly associate a particular strategy with a given function, so that a writer in choosing to create a text according to that scheme makes his function and strategy choice in one stroke: certainly, if he wishes to tell a story, he seems bound to use a narrative strategy. Yet he might also separate narrative strategy from its usual function and use it <sup>instead</sup> to reveal an entity or to express an idea.

\* Page 506.

The separability of function and strategy becomes clearer as one contrasts the descriptions of the functions, which are organic, pertaining to the whole text, with the descriptions of the strategies, which are analytic, pertaining to the nature and arrangement of the parts.

To identify how the text functions as a whole, the reader assesses his own response: what is he left with? For discourse function Type I, the entire text appears to have no other function than to "tell a story;" that is, to dramatize an event or sequence of acts/events that occurred in a single time frame. A change of some sort occurred over time, and this change is the focus of the discourse. In discourse function Type II, the text functions to reveal the nature of an entity (person, object, place, etc.) which retains its identity and essential properties over time. For discourse function Type III, the text reveals or expresses an idea or relationship, an abstraction from concrete events and entities (beyond the linguistic abstractions which merely serve to name them), a logical, analogical, or tautological construct that exists not in time but in mind. The same real world events can provide the given material for texts that function entirely differently. The role of Winston Churchill in World War II could be dramatized as pure story, the events related in a narrative sequence as entertaining and otherwise unedifying as any suspense tale; or the story could be told as a portrait of the man--rather, a personal history, with its emphasis on the hero's character, could be made from the same events. Finally, many sorts of analyses of the events in which Churchill figured could be made in support of a variety of assertions about cause and effect, the nature of war, the potential of one individual for historic impact, etc., creating Function III discourse about ideas or the meanings and interpretations of events or evaluations of people and things.

To some extent these differences in function appear at first to be simply a matter of choice of strategy: narrative, descriptive, or commentary. In fact, the strategies listed here are so widely used to express their parallel functions that a departure from convention is noteworthy, in some cases creating a distinct genre, in other cases a failed piece of writing. It is no wonder then that strategy and function have so often been combined or confused by theorists who sought a single basis on which to classify texts, or that theorists who classified according to function assumed that strategy naturally followed. This confusion may have resulted in part not only from similarities in the terms used to describe function and strategy, but from the use of the same terms to describe both whole strategies and individual propositions.

An adequate definition of strategy, or choice of text design, must distinguish between core propositions, forming the spine of the text, and other statements which elaborate, extend, provide background for these core propositions. Linguists have made this distinction for narratives or stories (Labov and Walletsky, 1966; Hopper, 1977); here I have extended it to other discourse types. I have often been disconcerted by efforts to describe texts as narrative <sup>versus</sup> / ~~descriptive~~ <sup>versus</sup> / ~~commentary~~ when I could find in almost every text a lavish sampling of all these kinds of writing. A second glance at most texts further reveals that the differences between them lie not merely in relative numbers of one kind of proposition or another. Rather, the key to text design strategies seems to lie in the relations of these different kinds of elements to one another.

Figure 3\* shows how a strategy can be identified by reference to which kind of propositions form the core of the text. In narrative strategy,

\* See page 507.

for instance, the core propositions, forming the spine of the text, are a series of statements, like, "X happened;" "X is happening;" or "X happens;" the latter (simple present tense) occurring only when it is clearly historical present, not habitual aspect:

Historical Present:  
(reference to one particular occurrence)

"The footsteps approach my door. I fix my eyes on the handle of the door, the gun steady in my hand. As the door opens, I shoot. A body crumples to its knees then topples into the room. With slowly dawning horror, I realize that I have shot my husband..."

Habitual Aspect  
(reference to a general, often repeated occurrence)

"For breakfast, I eat more than at any other meal of the day. I drink at least 8 oz. orange juice, and another 8 oz. milk, down 6 pancakes, 4 eggs, and finish up with toast or hash browns, with cheese and melon in season."

Although there may be flashbacks, or background description and commentary, the E statements at the core mirror the order of events as they occurred in real time, forming what Hopper (1980) calls the "foreground" of the narrative, with D and C statements providing "backgrounding."

E statements require transitive or intransitive verbs. Statements based on predicates which include linking verbs, the verb to be, or adverbs which indicate habitual aspect ("X usually is...;" "X sometimes does...;" "X used to happen;" "X would go and come..." (modal marks aspect in past tense)) are descriptive or D statements. These propositions appear as background in narration, but can be used to form the spine of a text, in which case the text strategy may be called descriptive. Note that these statements may be arranged to reflect a chronological sequence of events without creating a narrative text strategy: "Every day I get up, comb my hair, brush my teeth, yell at my little brother to make the beds, grab a bite from the fridge, and race off to the soccer field. Once there, I..." They may also be arranged to reflect spatial orderings in reality. Finally, they may be arranged associationally, mirroring the writer's thought processes and bouncing from one idea to another with connections visible only to the writer, or focally,

grouped according to some key aspect of the entity being described. (See Freedman and Pringle, 1982.) The term "entity" is used loosely here: such abstractions as "my summer vacation" may be treated as entities having consistent and identifiable characteristics, and may be written about using a descriptive strategy by presenting "a typical day in the course of..." It is a question in determining function, not strategy, to wonder just how abstract the subject of the piece is: when does summer vacation cease to function as an entity to be described and begin to function as an idea to be explored? Yet it is strategy that most often provides the answer to that question, the arrangement of propositions determining that a writer has crossed the line between descriptive and commentary.

In Type C strategy, common to exposition and argument, C propositions, or commentary, form the core of the text. E and D statements may appear in large numbers, with several E statements strung together to form a mini-narrative, but in all cases descriptive or narrative propositions will be subordinate to the C statements which they support, demonstrate, elaborate, define, etc. In addition, there may be C statements which are subordinate to the core propositions, that chain of assertions which provides the main thrust of the text. The core propositions can be arranged associationally or focally, as in Strategy B (description), or they can be arranged hierarchically, according to their logical relations. The assertions do not have chronological or spatial connections to one another, though they may be arranged to comment on chronological progressions and so appear chronologically ordered.

C statements are recognized in several ways: 1) the verb is marked by a modal: "X should, would, could, might, will happen...;" 2) the verb is marked by a negative: "X never does...;" "X did not happen...;" 3) the main proposition is imbedded as a noun clause of indirect quotation: "(Writer/X) believes, thinks,

feels, wishes, hopes that...;" or with a qualifying expletive: "It seems that...;" "It is clear that...;" 3) relationships of causality, comparison, opposition are expressed either through predication within a clause: "X is like...;" "X is caused by...;" "X causes...;" or by coordinating or subordinating conjunctions and introductory adverbs relating two clauses: "If X, then Y...;" "Because X, then Y...;" "Either X or Y...;" "Not X, but Y...;" "X is not true; rather, Y happened;" 4) the predicate defines and classifies the subject with relation to other things: "X is a kind of...;" "X is one of two kinds of...;" 5) the verb structure is like that of narration or description ("X is/was;" "X does/did ") but X represents an abstract subject: either a non-count noun ("money;" "water"), a hypothetical construct ("democracy;" "competence"), a feeling state or emotion ("love;" "anxiety"), a nominalized verb ("registration;" "confusion;" "obfuscation;" "segregation"), or a noun that has generic rather than specific reference, as shown by context:

<u>generic referent:</u>	→	<u>specific referent:</u>
" <u>the family</u> in America today is in danger of extinction."		" <u>The family</u> down the block is in danger of bankruptcy."
" <u>(The) people</u> who believe that kind of propaganda are uneducated."		" <u>The people</u> (who are) climbing into the lifeboats are unafraid."

Two kinds of development are involved as a writer masters these text design strategies: the acquisition of a new strategy, increasing the range of choices available to the writer; and the progression from novice to expert within one strategy. Both kinds of development almost certainly involve transition stages during which a writer may produce a text that is difficult to classify, or that is successful in neither one way nor another, representing only partial mastery of a text grammar. Figuring out where a student is on the road to full mastery of text design strategies is further complicated by the existence of sub-classes of each of the three strategies, as well as the possibility of deliberate mixes of strategy to accomplish certain ends.

Figure 4\* represents possible lines of development and acquisition for the three strategies described, and places these in the larger context of other discourse strategies not classified in this limited typology of student text types. The diagram distinguishes among three classes of strategies--pre-schema, open-schema, and closed-schema, which are probably acquired in that order by the language user in our culture. Pre-schema strategies are most comparable to Britton's "expressive writing," relatively idiosyncratic structuring of discourse reflecting the writer/speaker's own flow of thought or "inner speech," requiring little or no pre-planning, allowing the sender merely to think aloud or "compose at the point of utterance." (Britton, 1975.) I coined the term "pre-schema" to help understand papers in the sample that seemed strategy-less. On reflecting, I realized that following one's thought is itself a composing strategy, but one which does not recognize and use discourse schemas conventionalized in other people's discourse. To some extent, of course, what I have called pre-schema strategies have

become conventionalized, as fiction writers, particularly, present characters who narrate in their own voices. This kind of artful artlessness in the hands of skilled writers mimics the natural means of expression of writers/speakers who are not able or do not choose to adopt the strategies developed specifically to serve the needs of closed-schema discourse. In the student papers, it was generally easy to distinguish between spontaneous pre-schema strategies, usually appearing as rudimentary ruminations on strongly expository topics, and simulated pre-schema strategies in which a narrator other than the writer introduces him/herself before engaging in a self-exploratory interior monologue, a popular solution to the 1974 and 1975 topics about being someone or something other than yourself.

\* See page 508.

The open-schema strategies represented in Figure 4 were posited to account for another group of student papers (as well as literary prototypes) not classifiable as belonging to any of the closed-schema strategies. The distinction between open and closed discourse schemas is made by Bereiter & Scardamalia (1979) and others as an alternative to distinctions made between oral and written discourse. Typically oral speech occurs in open schemas, where one does not structure the discourse alone but is aided by conversational partners. Bereiter and Scardamalia point out that one of the difficulties for children learning to write is the problem of "going it alone," or creating monologues as opposed to dialogues. But not all oral speech is dialogue--as witness speeches or lengthy oral narratives; and not all written speech is monologue--consider active correspondences or note-passing in the classroom. Closed-schemas, discourses not dependent on interruptions and interactions, have features in common whether they involve oral or written mediums that distinguish them from open-schemas, which have their own features, present for both oral and written speech. Nonetheless, it is probably true that we learn open-schema strategies first because we learn oral speech first in its most common form, the oral dyad. Only later do we learn to make monologues, these becoming increasingly decontextualized or independent of shared speaker/listener context for their interpretation as we master the closed-schema strategies for supplying context within text.

Like pre-schema strategies, the open-schema strategies may be real, as when I write a real letter I intend to mail, or simulated, as when a writer tells a story in a letter or series of letters. The short story anthology Points of View (Moffett & McElheny, 1966) containing examples of simulated pre-schema discourse ("interior monologue" and journal entries) and open-schema ("dramatic monologue" and letters), was widely used in English composition

classes at Drake during the assessment years, a fact which may account for the appearance of these simulated open-schemas in the sample. In most cases, the elements of open-schema strategies (especially phatic comments, aimed at establishing a relationship between reader and writer rather than at communicating content: "HI! How have you been?" "Well, you're not going to believe this." "I know you'll think I'm crazy." ) are used to frame clear whole pieces of discourse using easily classified closed-schema strategies, in which case the writing was classified using the typology. In some cases, however, the student created such a strong sense that writer and reader were engaged in a dyad that the closed-schema typology could not apply--the papers belonged on another map.

In drawing the lines of development for Figure 4, I have attempted to show that expert use of any strategy is as cognitively advanced as expert use of any other strategy, and that writers may reach expert status for different strategies in almost any order, maybe becoming expert in only one strategy while remaining novice at all others. The direction of development, from left to right across the chart, is meant to suggest the likely order of initially acquiring the different strategies, given schooling in our society, and perhaps given the nature of cognitive development and the cognitive demands of the respective strategies. For instance, some aspects of commentative strategy appear to rely on a student's having entered Piaget's stage of "formal operations." Moffet (1968) describes the same direction of development, saying that young children write sustained concrete discourse (stories) with an occasional abstraction, and that only older children appear able to write sustained abstract discourse, learning to imbed the concrete references.

I anticipate a problem for evaluators when faced with the texts of children who are in transition between the points represented on the chart:

either because they are attempting to imitate a strategy they do not fully own, having just begun to learn it in school or having encountered it in reading, and do not yet have need of, not understanding the function for which the strategy was developed; or because they have encountered a new function--for instance, the need to express an idea--which their current strategy seems inadequate to deal with. This latter dilemma may drive a writer back into pre-schema strategy in a search for a language adequate to the new challenge. For evaluators concerned with documenting growth using improved scores on writing tests the problem is compounded when a student develops a high level of expertise in using one strategy, which he uses successfully in earlier writing, but later attempts to use another strategy at which he is a novice.

The writer has at least two other kinds of discourse options not represented in Figures 2-4, both of which may represent more cognitively complex task constructions than so far described because they involve re-combining elements of function and structure in less obvious ways. The first, already mentioned, are the cross-combinations of function and strategy made possible by separating these two aspects of composing.

A peculiar feature of the typology is that cross-combinations of function and strategy appear to be uni-directional: that is, more concrete strategies can be used to realize more abstract functions, but more abstract strategies cannot be used to realize more concrete functions. Figure 2 divides the discourse schemas that result from combinations of function and strategy into primary and secondary discourse types and provides the combining rule. It will be obvious to most readers that, while Function I is bound to narrative strategy, narrative strategy is not limited to the function of telling a story: a reader may complete a narrative realizing he has been preached to--that what was

at stake was an idea, not a set of events, a plot, a character. The possibility of using narrative to do more than tell a story may be at the heart of the difference between good second-rank entertainment novels in various genre and novels that compete as serious literature. This has nothing to do with what "lasts." Some good stories will last just because they are good stories. Other works will last because of something they reveal about the human condition, rather than because of a suspenseful plot and fast action.

One may argue, of course, that really good literature functions both as good story and as vehicle for an idea, which argument **introduces the second kind of re-combining of elements in more complex task constructions.**

I have found in the sample, as in the real world of adult writing, clearly distinct types of writing which seem to me to be best described as combining two functions, and/or two or more strategies. These mixes are not necessarily "mixed-up"; although inexperienced writers seem to lose track of function or shift strategy mid-way, producing mixed-up pieces, better writers seem able in many cases to produce controlled combinations identifiable as discourse schemes distinct from any of the pure types. Some examples will clarify the possible permutations of the system:

I Function: tell a story; with Stratgy A: narrative: One student wrote about the adventures of "Freddy the Fish" in response to the '75 topic; a picaresque tale with no apparent point, and no attempt to reveal why the writer might have wanted to share this particular life. Another student wrote what might have been a good episode for "Mission Impossible." Both used narrative strategy to tell a story, nothing more.

I-III Function: tell a story and reveal an idea; with Strategy A: narrative. A remarkable short story captures the moment of truth in the life of a young terrorist, who, feeling conscience-stricken in the act of setting charges to blow up a bridge, likens himself to the carnivores, who

by preying on deer and elk thin the herd and keep it healthy. This came in response to the 1974 topic, "choose to be something other than a human being":the student explores in his story the problem of why anyone might choose to be, or to see himself as, other than human.

I-III Function: tell a story and reveal an idea; with Strategy A-C: narrative and commentary. This class of writing is familiar as the autobiographical essay, or reminiscence, memoir, etc. which has before been inadequately characterized and confusingly classified by systems which distinguished narrative from expository writing and then had to decide where these pieces belonged. This kind of writing is widely featured in popular magazines, essay anthologies for Freshman composition courses, and as classroom assignments, from the first time a teacher says, "Write about your summer vacation." Unlike good story-telling which requires that the commentary serve the narration, or good exposition, which requires that narration be subordinated to commentary, the personal experience essay requires a balance of both strategies. It is organized and functions as a story, but it is the story of an idea--a realization, a discovery, an insight the writer had as a result of a series of events. The discussion of the ideas--what I understood before, what I understand now--is as important as the relating of events, but is itself couched in a narrative, or at least a chronological framework. In an effectively fused piece, a reader might find it difficult on concluding to decide whether she had just read an essay or a short story--simply because both those terms must be used so loosely to apply to this kind of contemplative autobiography.

III Function: reveal an idea; with Strategy C. This kind of text was common in response to the 1976 topic, naturally, and represents traditional notions of formal argument or exposition. A thesis is set forth, and is followed by a series of commentary propositions, with any descriptive or

narrative material introduced as elaboration or support, a kind of backgrounding for these core comments which make up the writer's argument.

III Function: express an idea; with Strategy A. There was only one clear case of this in the sample, but prototypes in literature would include morality tales and allegories, where the concrete events in the story represent what are believed to be general truths, not real occurrences.

Combinations of Function II and Strategy B with other functions and strategies will be fully described in the report on the Drake data. At first count after coding, it appears that the typology was able to account for better than 90% of papers in the sample, with the other 10% falling into pre-schema or open-schema categories. The primary types, with matched function and strategy, appeared to be marginally preferred solutions for every topic/year, although which primary type was preferred sometimes differed within topic, depending on grade level. Almost every kind of mix listed in Figure 2 as possible was included in the sample, with some mixes causing more trouble in coding than others. No attempt was made in the early stages of coding to distinguish between "mixed-up" papers and apparently controlled combinations; it is expected that different kinds of mixes can be coded in the future to provide additional information about a writer's level of control over his or her chosen strategy or discourse type.

#### CHOOSING A STRATEGY

In some cases, the success of a piece of writing in which a concrete strategy fulfills an abstract function may be accidental. That is, the writer may be unaware of the rich and complex meanings readers are able to infer from his text. In the Drake sample, however, successful crossing of strategy and function are generally regarded as intentional and are seen to represent a highly complex task construction, available only to more experienced composers.

Similarly, a successful mix of two functions or two strategies tends to result in a text that is recognizably like a prototype in adult prose, so that the writer can be credited with having attempted a more complex task, the creation of a secondary discourse type. But when one of these conventional kinds of mix is not in evidence, it is not clear whether the writer was attempting such a mix or was merely shifting ground, attempting to cope with topic constraints she felt were conflicting, or to cover weaknesses in her own repertoire of strategies.

Empirical research is needed to help discover why a student construes a task as he does: what is the range of options he thinks he has? How does he want his text to function and why did he set that goal? What strategies will he choose to reach his goal? Is he an oldtimer or a newcomer to his chosen strategy? What awarenesses of text grammar appear to influence his whole text planning or his sentence level planning? How is the onset of text level planning related to increasing awareness of discourse schemas? The discourse typology, refined and clarified, may be of some use in exploring these questions, questions which may help us discover why and how students change the difficulty of what they attempt, even when a tester may intend task demands to be parallel. These questions may also help us define task complexity in new ways, allowing more precise evaluation of changes or growth in composing ability.

While discourse functions may be recognized on a deep, fairly unconscious level, I believe that strategies can be made conscious, learned deliberately, and applied selectively. On the other hand, experience with writers like those in the sample suggests that students, in selecting a strategy, are generally working from a confusing array of instincts and inhibitions which have little to do with the appropriateness of the strategy to the desired function.

Table 2 outlines some factors I think may influence a student's choice of strategy, whether pre-schema or schema-based. Future research will determine whether these are real and how they may operate for given students. I may assert, logically, however, that availability is absolutely determining. A writer cannot use a strategy which is not available to him. Given the availability of more than one schema-based strategy, together with the universal availability of the strategy I call pre-schema, any or all of the remaining factors may come into play in any combination. I suspect that some of these create conflicts for the writer and that keeping these factors in mind will help researchers understand papers whose task constructions are not easily classifiable by the typology.

It is premature to judge how well this typology will assist efforts to define task complexity in writing or to assess a student writer's level of development on the basis of texts produced for evaluation. So far, informal sharing with teachers of composition suggests that the typology can have a clarifying effect of distinguishing among types of writing tasks set by or for students and can help teachers identify some of the strategies students may have only partially mastered or may be moving between. A teacher of high school juniors identified two kinds of text typical of students in intermediate stages between mastery of narrative and mastery of commentative strategy. She submitted these two samples:

Assignment: It is said that we learn best from our mistakes. Agree or disagree, drawing on examples from your own experience.

Stage One, student response: "It is said that we learn best from our mistakes. I know of one time when I learned an important lesson by doing something stupid. I was twelve at the time... (narrates event: four pages of lively story-telling.)  
...I thought my mother was paranoid, but now I know I'll be telling my kids, "You shouldn't trust strangers."

Stage Two, student response: "It is said that we learn best from our mistakes. That is probably true because until we make a mistake we don't have any reason to want to know something. I have learned a lot of things by making mistakes, and most of these have stayed with me more than the things I learned by studying courses in school or reading or listening to what my parents or other people told me. Maybe those things stuck in my mind because it's so humiliating to make a mistake. You can't forget the embarrassment, and you'd rather risk death than risk that again.

"It's good to remember that our mistakes teach us something and that we don't have to be ashamed to do something wrong, especially when we didn't know any better. That is how we learn. Being afraid of making mistakes can keep you from learning."

The first student is still largely committed to narrative strategy: her introduction and conclusion are concessions to the expository nature of the prompt, but they are minimal, interfering as little as possible with the story. Students less aware of essay requirements often ignore even these concessions, beginning, "It happened when I was twelve..." or, "One time, when I was twelve, ..." A teacher wanting to lead a stage one writer further into the language of ideas might ask the student simply to expand the commentary of the opening and/or the closing, or to write a second narrative of another time when she learned from a mistake, then to write a paragraph comparing or contrasting the two experiences, drawing some conclusions about how or why people learn from mistakes.

The second student has fully abandoned narrative strategy, with a common result: he is able to produce only two paragraphs. These paragraphs are rich in insights worth exploring in an essay. Further, they present a paradox which will require more thinking by the student if he is to untangle it: the first paragraph suggests that being afraid of making mistakes, or being ashamed, is why we learn from our mistakes; the second suggests that being afraid of making mistakes may keep us from learning. Two good ideas; the paradox is only apparent. But the paradox is apparent, rather than hidden,

because the student risked talking about ideas, struggling to make explicit understandings which were only intuitive before. The struggle is essential if the student is to acquire that all-purpose academic discourse schema: Function III, Strategy C, expressing an idea using commentary as the core of the text, working out the logical relations between the parts of the idea. But the stage two student is often greeted with such negative evaluations he may be excused for wanting to retreat to the old successes of stage one or even pure narrative. "This is too short. You contradict yourself. You don't have a single concrete example from your own experience. You shift points of view, from first to second person; and what is this? the royal 'we'?" As the student says, "Being afraid of making mistakes can keep you from learning."

Neither of these essays would receive a top score during a holistic assessment in which at least some students managed to produce Strategy C essays of better than novice quality, essays which included a full range of propositions from abstract to concrete, with concrete description or narration carefully attached to commentary statements, which in turn were arranged to form a coherent core argument. But it is highly likely that raters would prefer the stage one paper with its expert use of narrative strategy to the stage two paper of the novice commentator with its all too evident flaws. No provision could be made for recording that the stage two writer may have moved closer to the goal than the stage one writer if one compared their respective task constructions.

Much greater clarification of discourse features, and of the role played by text level expectations of both writers and readers, seems needed before it will be possible meaningfully to measure improvement in composing abilities.

## NOTES

- 1 The papers in the sample were pooled across all topics, all years, in a single holistic scoring, to obtain scores that allowed comparison of all papers according to a general impression of quality. Details of this scoring will be available in the final report on this research, from the NIE Writing Assessment Project, Marcia Farr, NIE Project Officer, Reading and Language Studies, Grant No. NIE-G-80-0034, c/o the Bay Area Writing Project, University of California, Berkeley. Working title: Technical Report No. 5, The Drake Longitudinal Study.
- 2 It is important not to confuse task complexity with text complexity. Joe Williams makes this critical distinction (1979), pointing out that very complex cognitive activity may be required to produce a text that is simple and direct from the reader's point of view, while a very complex text may be produced by the simple expedient of transcribing thoughts as they occur to the writer.
- 3 Kinneavy (1979) describes the aims of discourse in a volume which exhaustively analyzes examples of each type, attempting to establish stylistic features these texts have in common, while reserving his discussion of the "modes" of discourse to a later volume, not yet printed. It is with some hesitation that I offer yet another discourse classification, rather than waiting for his definitive second work, but I find his system useless for the student sample in question, as well as inadequate to account for many text types I have encountered in adult letters, both in and out of academia. The separability of function and strategy suggests the need to classify each text on both counts in order to describe its essential differences from other texts, rather than first classifying all texts according to function or aim, as Kinneavy appears to do, reserving til later an attempt to describe their modes. As can be seen in the application of my typology, Function A in combination with Strategy A results in a very different kind of text from a text which combines Function C with Strategy A, and this difference establishes a different class of text rather than one being a sub-species of the other.

## REFERENCES

- Bain, A. English composition and rhetoric. American Edition, Revised. New York: D. Appleton & Company, 1890.
- Bereiter, C. & Scardamalia, M. From conversation to composition: the role of instruction in a developmental process. To appear in R. Glaser (Ed.), Advances in instructional psychology. Vol. 2. Hillsdale, N.J.: Erlbaum. (Ms., draft: November 6, 1979.)
- Bever, T.G. Regression in the service of development. To appear in T.G. Bever, et al. Regressions in development: basic mechanisms and phenomena, in press. (Ms., draft: 1981.)
- Breland, H.M., & Gaynor, J.L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, summer 1979, 16 (2), 41-49.
- Britton, J. Language and learning. Urbana, Illinois: National Council of Teachers of English, 1975.
- D'Angelo, F.J. Modes of discourse. In G. Tate (Ed.), Teaching Composition: 10 bibliographical essays. Fort Worth, Texas: Texas Christian University Press, 1976. 111-135.
- Diederich, P. Measuring growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.
- Freedman, A., & Pringle, I. Writing in the college years: some indices of growth. College Composition and Communication, XXXI: Oct. 1980. 311-324.
- Godschalk, F.I., Swineford, F. & Coffman, W.S. The measurement of writing ability. New York: College Entrance Examination Board, 1966.
- Goodman, D.R. Stage transitions and the developmental trace of constructive operators: an investigation of a neo-Piagetian theory of growth. (Doctoral dissertation, York University, Toronto, Ontario, 1979. Investigates the problem of uneven development applying the theories of Juan Pascual-Leone.)
- Hopper, P. Aspect and foregrounding in discourse. Paper presented at the Symposium on Discourse and Syntax, UCLA, November, 1977. Unpublished ms. available from author, Linguistics Program, State University of New York, Binghamton.
- Keech, C., and McNelly, M.E. Comparison of student and teacher rater responses during holistic scoring: differences in task interpretation. In C. Keech, et al. Effects of variation in a writing test prompt on holistic scores and composing process. (Part Two.) Technical Report No. 2 of the NIE Writing Assessment Project, Marcia Farr, NIE Project Officer, Reading and Language Studies. Grant No. NIE-G-80-0034. Berkeley, California: Bay Area Writing Project, 1982.

- Kinneavy, J. L. A theory of discourse: the aims of discourse. Englewood Cliffs, N.J.: Prentice Hall, 1971.
- Labov, W. & Walletsky, J. Narrative analysis: oral versions of personal experience. In J. Helm (Ed.). Essays on the verbal and visual arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society. Seattle: distributed by the University of Washington Press. 1966.
- Meyer, B.J.F. The organization of prose and its effects on memory. Amsterdam: North-Holland Publishing Company, 1975.
- Miller, S. Rhetorical maturity: definition and development. Reinventing the rhetorical tradition. A Freedman and I. Pringle (Eds.), University of Central Arkansas: L & S Books, 1980, 119-127.
- Moffett, J. Integrity in the teaching of writing. Phi Delta Kappan, December 1979, 276-279.
- Murphy, S., Kinzer, C., & Carroll, K. A study of the construction of the meanings of a writing prompt by its authors, the student writers, and the raters. Technical Report No. 4 of the NIE Writing Assessment Project, Marcia Farr, NIE Project Officer, Reading and Language Studies. Grant No. NIE-G-80-0034. Berkeley, California: Bay Area Writing Project, 1982.
- Piaget, J. The development of thought: equilibration of cognitive structures. New York: The Viking Press, 1977.
- Spandel, V. & Stiggins, R. Direct measures of writing ability: issues and applications. A report of the Clearinghouse for Applied Performance Testing. Portland, Oregon: Northwest Regional Educational Laboratory, 1980.
- Stein, N. & Glenn, C.G. An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.). New directions in discourse processing (Vol. 2). Norwood, N. J.: Ablex Publishing Corporation, 1979.
- Williams, J. Defining Complexity. College English, Vol. 40, No. 6: Feb. 1979.

#### Addendum

- Freedman, A. & Pringle, I. Why children can't write arguments. (Linguistics department, Carleton University, Ottawa, Ontario.) Unpublished ms. 1981.
- Moffett, J. Teaching the Universe of Discourse. Boston: Houghton Mifflin, 1968.
- Moffett, J. & McElheny, K. Points of View. New York: New American Library, 1966.

Table 1

Topics Used in Six Years of Drake Writing Assessment

- 1973 Write about an event you wish you had witnessed or could witness. The event can be real or imagined; the time of the event can be past, present or future. Make it clear why the event is significant to you. You may write a journal entry, letter, dialogue, monologue, essay, story, autobiography, or other form.
- 1974 If you had to choose to be something other than a human being, what plant or animal or other form would you choose? In your writing, give your reader some idea of what you think it would be like to be that form, and of why you chose it. You may do this writing as a journal entry, a letter, a dialogue, a story, an autobiography, an essay, a poem, etc.
- 1975 If you could change places with someone else, who would it be? The person you write about can be living, dead, drawn from past or present, from books, films, etc., or from your own imagination. In your writing give your reader some idea of what it would be like to be that person, and of why that life appeals to you. You may do this writing as a journal entry, character sketch, dialogue, letter, story, autobiographical essay, argument, poem, or other form.
- 1976 Not all inventions have been good for all humanity. Name one invention we would be better off without, and make it clear why. You may do this writing as an essay, journal, letter, story, or other form.
- 1977 Imagine that a small group of people will be sent to colonize a new planet. Food, clothing, shelter and transportation have been provided for. You are among those asked to select a few additional things to be sent along in the limited space available in the ship. What one item would you recommend, and why? You may write your recommendation in the form of a story, a dialogue, a letter, a speech, an essay or other form.
- 1978 Write about some way in which your life has been, or might be influenced. You might write about the influence of another person, a book or film, an idea, or an event such as a triumph or defeat, or a sudden gain or loss. Make it clear just what or who influenced you, and what the effect was upon you, or what the effect could be upon you. You may do this writing as a journal entry, character sketch, dialogue, letter, story, autobiographical essay, or other form.

## Table 2. CHOOSING A STRATEGY

A list of factors which may determine or affect a student's choice of strategy, whether pre-schema or schema-based.

1. AVAILABILITY. (Which strategies does the student have in repertoire?)
2. PREFERENCE. (Does the student prefer some strategies because of familiarity, facility, regardless of purpose or context of the writing task?)
3. APPROPRIATENESS TO CONTEXT. (Does the student have ideas about what strategies are appropriate to writing tests, regardless of topic, or the importance of applying recently learned strategies to demonstrate mastery, regardless of their naturalness for a given purpose or assignment?)
4. APPROPRIATENESS TO GIVEN TASK. (Certain topics suggest explicitly, or tend to elicit certain strategies, by eliciting certain purposes which are most easily accomplished with matching strategies. e.g., tell a story is only accomplished by narrative.)
5. APPROPRIATENESS TO WRITER'S CHOSEN PURPOSE. (If writer chooses a purpose other than the obvious one suggested by the topic, he may choose a strategy appropriate to his own purpose rather than one appropriate to the expected purpose suggested by the topic.)
6. IDIOSYNCRATIC OR HIGHER ORDER CRITERIA. (The writer may have personal criteria, such as wishing to be novel or unusual, which may influence him to choose an uncommon solution: for instance, deliberately mismatching purpose and strategy because he does not wish to do the obvious, or because he believes he can achieve his purpose more effectively, perceiving a complex interaction between form and function which will communicate more than he might should he use expected strategies for a given purpose.)

*C. L. Keech*

	1973	1974	1975	1976	1977	1978
FRESHMAN	Cohort I 30	Cohort II 30	Cohort III 30	30 papers	30 papers	30 papers
SOPHOMORE	30 papers	Cohort I 30	Cohort II 30	Cohort III 30	30 papers	30 papers
JUNIOR	30 papers	30 papers	Cohort I 30	Cohort II 30	Cohort III 30	30 papers
SENIOR	30 papers	30 papers	30 papers	Cohort I 30	Cohort II 30	Cohort III 30

30 students 30 students 30 students  
 120 papers 120 papers 120 papers  
 Cohort I, Cohort II, Cohort III,  
 class of 76 class of 77 Class of 78

Total papers = 720  
 Three longitudinal samples = 360 papers  
 Supplementary sampling = 360

Figure 1. Sample Population

Figure 2.

A Discourse Typology: describing the task constructions found in student prose written in response to 6 prompts.

	<u>Discourse Function</u>	<u>Discourse Focus</u>	<u>Discourse Strategy</u>
↑-----concrete----- ↓-----abstract-----	I. TELL A STORY	Change over Time	A. narration
	II. REVEAL AN ENTITY	Identity over Time	B. description
	III. EXPRESS AN IDEA	Relations beyond Time	C. commentary

Discourse types, resulting from combinations of function and strategy found in the sample or in adult literature:

1. Primary types, matched function/strategy: IA; IIB; IIIC.
2. Primary types, mixed function/strategy: IIA; \* ; IIIA; IIIB.
3. Secondary types, matched function/strategy: I-III/A-C; I-II/A-B; II-III/B-C.
4. Secondary types, mixed function/strategy: I-II/A; II/A-B; I-III/A; I-III/A-B; II-III/A; II-III/A-B; II-III/A-C; etc.\*

\*Combining rule (empirically derived) allows more concrete strategies to serve more abstract functions, but concrete functions cannot be served by more abstract strategies: hence, Function I can only be realized by Strategy A; Function II by strategies A and B (ideally); and Function III by any strategy, A, B, or C. By this rule there should be no IB, or IC (true, in fact); or II/C, or II-III/C (these latter occur.)

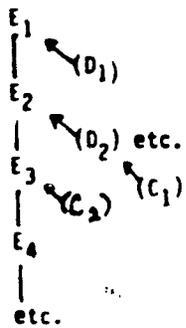
C. Keech

521

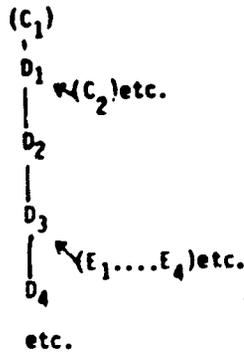
Figure 3.

Proposition Structure of Three Discourse Strategies:

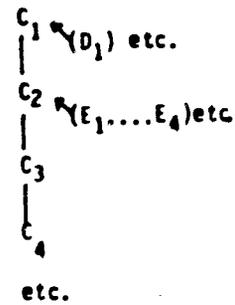
NARRATIVE



DESCRIPTION



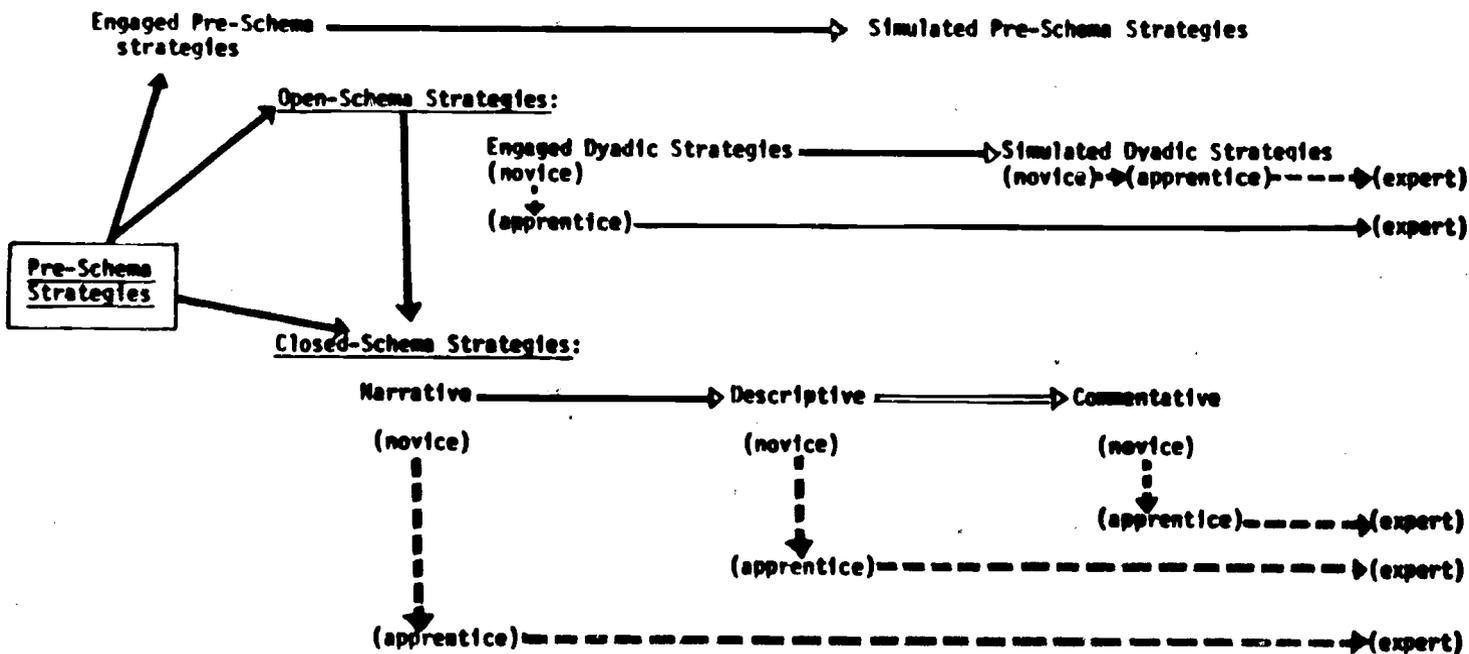
COMMENTARY



- E = Event statements: "X happened." "X is happening." "X did..."
- D = Descriptive statements: "X is..." "X does..." "X happens..." (habitual aspect and/or linking verbs in predicate; subject of verb refers to actual entity or action treated as entity, not to hypothetical construct.)
- C = Commentative statements: (assertion, interpretation, evaluation, classification, etc)  
 "X should be..." "X might be..." "X did not..." (modals, negatives)  
 "X is caused by..." "X causes..." "X can be classified as..." (relatives)  
 "X means..." "That X happened signifies..." (interpretations)  
 (Also, descriptive statements like "X is..." become commentative when subject or subject complements are hypothetical constructs.)

C. Keach

Figure 4. Hypothesized Chronology of the Acquisition and Development of Writing Strategies.



→ hypothesized direction of development, between schema types.

⇒ hypothesized direction of development within schema types, between strategy types

- - - → logically required direction of development within strategy, which may in actuality be broken by stops and reversals along the way

C. Keach

BEYOND HOLISTIC SCORING: RHETORICAL FLAWS  
THAT SIGNAL ADVANCE IN DEVELOPING WRITERS

Elissa Warantz and Catharine Keech

Introduction

The study analyzes a longitudinal sample of student exams collected across four years from 90 high school students who participated in annual writing assessments. The writing test each year consisted of an hour-long composition written to a prompt which allowed the students to draw on personal experience or imagination and to choose their mode of discourse. All of the essays were holistically re-scored in a single pooled sample for purposes of this study. Holistic scores represent single-impression rapid estimates of over-all rhetorical effectiveness of an essay compared to other essays in the sample. A conspicuous problem recurring consistently in this widely used assessment procedure is the instability of student performance from one essay test to another, as exemplified by low test/re-test correlations, by failure of individuals and groups to produce consistent improvement on these essay tests from year to year even with concentrated writing instruction, and by the frequency with which individual or group mean scores actually decline on post tests or second trials.

Attempts to solve this measurement problem in essay testing have focused on developing more precisely defined topics and on more reliable rating procedures, but there have been no sustained efforts to explain what actually changes in student writing across several test occasions. This study investigates the longitudinal dimension of this phenomenon by following student performance across four annual test occasions. The study shows that some students attain lower holistic ratings on later test essays because they use more complex or difficult composing skills or because they have defined the rhetorical task in the assigned prompt more complexly than in a prior test. Since the loosely defined writing prompts offer ample opportunity for individual variation in task interpretation, students who have learned new skills or who begin to perceive writing tasks more complexly may attempt more than they can do easily, with the result that their essays are less rhetorically well-formed than on earlier occasions when they attempted less.

This apparent regression in performance which actually

signals movement forward in cognitive or affective development or learning has been termed "disequilibrium" by Piaget in his stage model of cognitive development

### Purpose of the Study

The purpose of this study is to examine the extent to which students attempt to do different kinds of writing, either spontaneously or in response to strongly suggestive cues of the writing prompts they receive in a series of annual holistic writing tests taken during their ninth to twelfth grade years in high school. It was hypothesized that differences in the student's task construction, or what the student attempted to do, might help to account for unexpected directions of change in student performance scores from year to year. In other words, when a student does the same kind of writing in 1976 that he did in 1975, using similar strategies to accomplish similar purposes, an evaluator might reasonably compare the student's two scores to determine whether he was doing the same thing any better after a year's maturation and some writing instruction. If, on the other hand, the student does something different in 1976, using new strategies to accomplish new purposes, his score will not indicate how much he has improved at doing the old task; rather, it will reflect how well he does on an altogether new kind of task. This research is a first stage effort to distinguish among the factors which contribute to the seeming instability of student performance from one test occasion to another. Ultimately, this research will direct attention to the developmental features of text design and task analysis which also influence quality rating of student essays in seemingly paradoxical ways.

The improvement of quality ratings of student performance on writing assignments has long been equated with improving student writing abilities, and this advancement in quality is generally accepted as the goal of writing instruction. Even when educators use multiple choice tests as indirect measures of writing skill (because of their ease of administration and the superior score reliability of these measures) they recognize that such tests must ultimately be validated by relating their scores to performance on direct measures, actual writing tasks in which the student demonstrates his ability to construct or compose a response, not merely select a response (Godschalk, 1966; Diederich, 1974; Breland and Gaynor, 1979; Cooper and Odell, 1979; Spandel and Stiggins, 1980; Schools Council, UK, 1966; NCTE 1975).

As noted above, a perplexing problem has baffled and frustrated teachers, researchers, and evaluators of writing instruction. For as children learn to write, they should perform better: their scores on successive essay tests should improve. If children don't increase their scores, it is easy to assume that they have made no advances in writing

ability. Alternatively, one can seek other explanations. One may examine the test measures used: were the two test instruments truly parallel in what they required students to do; were the two tests scored in the same manner? Testers have reported improving score reliabilities substantially by improving inter-rater agreement, by phrasing prompts carefully to be sure demands are parallel, by taking longer writing samples to improve discrimination of the measure, and by taking more than one writing sample at each test sitting (Breland and Gaynor, 1979; Steel, 1979). All of these procedures help reduce the measurement problem to some extent; however, none examines what actually changes in student writing from test to test; therefore, these corrective measures fail to reveal a third likely source of variation in student performance which has radical implications for both evaluating and teaching composition skills.

To understand variations in student writing performance on tests it is important to distinguish between the examiner's given task-- the set of constraints and options expressed or implied in the text of the writing prompt-- and the student's construed or constructed task-- the set of constraints actually honored (and the options actually chosen) by the student in composing. The student's constructed task is derivable from the text of his composition by noting the task constraints honored. While test makers do attempt to control the given task to create parallel task demands on alternate occasions, they are less able to control the student's construal of the task.

In other performance areas like music, a learner is given a more difficult piece to play or a new instrument to learn, and in mathematics, both teacher and student can be aware of degrees of difficulty in problems. But in composition it is possible to give students identical tasks on two different occasions and have them construct such different composing tasks that in an early version of the task, a student might well appear expert in his response to the problem he set himself; while in a later version, he or she might suddenly become a novice as he or she attempts to solve a far more complex problem. That is, writing prompts are usually loosely structured in the sense that the particular given task (writing assignment) may be construed as offering a range of possible real tasks, any of which the writer may elect to do.

In an effort to rank writing assignments according to difficulty level, it is not sufficient just to compare the literal texts of the prompts, which ostensibly present a particular given or intended task. Students will inevitably vary in how they interpret and respond to that text and in how they construct the task that they will actually attempt (Murphy et al., 1982; Keech, 1982). Open-ended assignments, simply stated, may be offered again and again over all the years of a writer's career from his first awkward printings

to his adult years of proficient composing; the actual difficulty of the assignment, however, will increase in proportion to the student's very ability to define the task more complexly. Reasonable task definition is one of his most important writing skills, and will develop along with syntax, sense of paragraphing, and other linguistic knowledge. An unreasonable narrowing of topic constraints on the other hand, may not only inhibit variation in construal, but it also may limit measurement of the student's ability to juggle constraints, set goals, make plans, select strategies--all composing skills as important to measure as the student's ability to transcribe language using correct written conventions. In other words, determining the student's individual construction of the task is of primary interest to evaluators and researchers who wish to measure a student's growth or improvement in writing ability.

What is needed, as a preliminary to any effort at measuring development of writing ability is a method of comparing what students actually do and how they do it, from one occasion to the next. A system of task descriptions is required, through which student texts can be analyzed to determine whether or not two compositions represent the same or different types of tasks: which purposes does the student appear to be aiming for? which strategies does he use? Only after we are able to describe more exactly what the student actually produces in response to test prompts, will we be able to infer the student's conceptualization or construction of the writing task. Then we will be able to generate and test hypotheses about the relative level of difficulty of what is attempted on different occasions. That research should, in turn, make possible a proper investigation of the nature of developing writing skills, by allowing us to distinguish between a student's mastery of old forms long practiced, and his novice attempts to try out new patterns. Further, such research may ultimately allow researchers to predict which kinds of tasks are learned first, which follow, and what various spontaneous task constructions employed by students may indicate about a given student's particular level of development.

### Procedures of the Study

The first step in this research, therefore, was to derive a typology of task construction which would uniquely characterize the range and variation of the written texts produced by students in the longitudinal sample as well as the kinds of texts traditionally classified by discourse theorists (Bain, 1890; Kinneavy, 1979; Britton, 19\_\_; Emig, 19\_\_; etc.) in terms of literary "genres." Once it was derived, the typology was applied to a sample of student papers, and cross-tabulations were drawn to determine concomitant variation in prompt, grade level, task construction, quality rating (holistic score). This study thus provides us with data through which we can compare

Individual student performance (in terms of task construction and score) to the favored task construction and mean score of all students in a given topic/year. Students who appeared to deviate in either or both dimensions from the developmental pattern common to the larger sample, as well as students who appear best to typify the common pattern, will be selected for detailed case study in future research.

**The Data Base.** A longitudinal sample of student essays was drawn from the total set of essays produced during annual writing assessments at Sir Frances Drake High School, during the years 1973 through 1978. The sample selected for the study consists of three sub-samples of longitudinal data and one sub-sample of cross-sectional data, as shown in **Figure 1.**

Three cohorts of students were identified: a cohort consists of 30 students who attended Drake High School together, completing all four years, from freshman year through graduation. Cohort I attended DHS from 1973 through 1976 (actually entered in the fall of 1972) and participated in four spring writing assessments; Cohort II participated in the writing assessments conducted from 1974 through 1977; Cohort III participated in writing assessments in 1975 through 1978. The thirty students forming each cohort were drawn from a larger pool of all students who completed four years at Drake and for whom test papers were available. The original sample was stratified for senior performance only, into high, middle and low scoring writers. Roughly proportional random samples were drawn from the three scoring strata, slightly larger than the final sample of 30 students to be used in making the final cohorts. The selected seniors were traced through preceding years, with cases dropped from the cohort if they did not produce a total of four papers over the four years. With slightly more than 30 cases still remaining in each cohort, it was possible to drop additional cases using sex as a criteria, so that

approximately equal numbers of males and females remained in the final cohorts. The resulting sample consisted of three longitudinal sub-samples, or cohorts of students, with 30 students in each cohort. Each student had produced four essays over four years of writing tests.

In addition to the longitudinal sub-samples, a cross-sectional sample was drawn from the original set of essays to represent student performance on each assessment topic, within grade levels not represented in the longitudinal cohorts. For instance, the longitudinal sample produced only freshmen papers for the year 1973. In analyzing these papers, it would be misleading to attribute certain features of the writing to the fact that the writers were freshmen if we did not also have upper-class student writing to check whether these features were particular to responses to the 1973 topic, rather than simply a result of freshman writing tendencies. The cross-sectional "simulated cohorts" provide a normative matrix for topics, as well as writing produced by any given grade level in response to several topics.

A different writing prompt was offered each year; topic and year may then be combined in a single variable called topic/year, indicating that identifying a particular year, e.g., 1973, also identifies a particular topic which was not repeated.

In the six assessment prompts used at Drake, the writing instructions remained fairly similar from year to year, allowing students to choose among a variety of genre, such as story, essay, journal, etc. What changed from year to year was the topic, or content focus. Thus, patterns of differences in student responses from year to year appear to be a result of differences in the way the test question described the actual topics of the writing, rather than differences in instructions to the writer about how to respond (what genre, or strategy, to use). Hence, a major variable in the study is topic/year, or the unique topic offered in each assessment year. These are identified in this report as follows:

- 1973: The [EVENT] topic
- 1974: The [BE] a [NON-HUMAN] thing topic
- 1975: The [BE] ANOTHER [PERSON] topic
- 1976: The [INVENTION] topic
- 1977: The [SPACE OBJECT] topic
- 1978: The [INFLUENCE] topic

The complete prompts are included in the Appendix as Table 1, and are typical of writing assignments offered in schools, particularly for the purposes of making over-all impressionistic quality ratings of the kind used in the Drake writing assessment program.

Two impressionistic quality ratings exist for each essay: an original rating (Old Score) assigned to the essay by teachers at Drake the year of the test administration, in which the essay is compared to all others written on the same topic, including essays by students at all grade levels, i.e., by the full school population at the time of testing. A second quality rating (New Score) was assigned to each essay in the sample during a pooled re-scoring of the essays, in which the essay is compared to all others in the sample (see Table 2 in the Appendix).

**The Student Writers.** The population of students attending Sir Frances Drake High School, in Marin County, California, remained reasonably stable during the years encompassed by the study and consisted of middle to upper-middle class students, with few students who could be characterized as educationally handicapped, non-standard dialect, or English-as-a-second-language students. Emotionally handicapped or learning disabled students (cases of dyslexia or other perceptual disabilities) were not identified and isolated or removed from the population sampled, either before or after testing. On the average, a considerable percentage of Drake students go on to two-year or four-year colleges.

**The School Writing Program.** During the years of the assessment program, from 1967 through the present, English teachers at Drake must be described as exceptionally able teachers of writing. The core staff in the early years authored a series of composition teaching guides for Ginn (Lavin, et al, 1965) which are still widely popular. The same group was active in leadership roles in the California Association of Teachers of English and provided the largest number of participants from one school in the establishment and initial sessions of the Bay Area Writing Project. The assessment program at Drake was introduced by Kate Blickhahn, then department chair, based on the model of so-called holistic scoring used by the Educational Testing Service, and later disseminated through the Bay Area Writing Project to local school sites throughout the country. Although the Drake teaching practices may not yet be typical of writing teaching in the country as a whole (Applebee, 1980), they may represent a vanguard approach to writing instruction, which helps to assure the author that insights into student writing development gained in the present study will not quickly become irrelevant as improved instruction in writing changes the nature and course of writing development in our culture.

**The Impressionistic Quality Rating or "Holistic Score"**  
The general procedures for obtaining quality ratings, hereafter called holistic scores following usage in the profession, were similar for both old scores and new scores, and are described in detail elsewhere by Keech (1981a). In summary, the procedure involves rapid-impression rating of

student essays by two independent raters, each using the same range of scoring points (in this instance, a scale of 1 to 9 with 1 awarded to the poorest papers and 9 awarded to the best papers in the sample). Inter-rater reliability is improved by pre-training using a scoring guide and sample papers, or "anchors," which illustrate typical qualities of writing to be found at each point in the scoring range.

Readers in the pooled re-scoring of the sample for this study included teachers who had read in the original Drake assessment, as well as additional BAWP teacher consultants experienced with the Drake approach to holistic scoring. The holistic scores thus obtained, both old and new, are representative of conventional and traditional impressionistic quality ratings of student work which have been treated as dependent variables in numerous evaluations of writing instruction, and have themselves been subjected to repeated analysis by researchers attempting to identify the factors which influence raters' judgments during these and similar scorings. (Freedman, 1981; Freedman and Nold, 1977; Haswell, 1980; Diederich, 1974).

### Selection of Cases for Analysis

The three longitudinal cohorts provided a total of 90 potential case studies of students performing on writing tasks over four years. Since exhaustive analysis and coding of all 90 cases was not possible, ten to twelve cases were selected from each cohort (or one third of the total sample) for case-study treatment which will provide the bulk of the findings reported in this study.

The primary criterion for selection of cases of interest was the appearance of a non-chronological sequence of performance quality ratings, as indicated by an Adjusted Ranking. This ranking combines the rank-ordering of the original holistic score (OS, in which the paper was compared to all other papers on the given prompt), the new holistic score (NS, in which the paper was compared by the same raters to papers on all prompts), and a forced-choice ranking (FC, in which 2 independent raters compared the papers of one student only to other papers by that same student. The combined rankings allow an adjusted ranking (AR) which reflects all 3 fields of comparison and, thus, all of the criteria likely to be used by teachers or raters in evaluating student samples to determine improvement in writing ability. Ideally, the AR was designed to remove misleading effects of different comparison fields, and to produce a ranking which agreed with the chronological order in which the papers were produced. The case whose rankings were not "correctable" by the AR were those chosen for study; i.e. student papers whose AR did not conform to the expected chronological direction of change (one to four, lowest to highest AR, freshman to senior) were chosen as case studies.

Originally, we had also intended to select case studies according to a second criteria, which involved comparison of a student's four new scores with a mean NS in that student's cohort during the four topic/years. It seemed clear that topic effect was reflected in mean new scores for each cohort (e.g. the mean NS for Cohort I rises dramatically between freshman and sophomore years-- $+2.44$  points--by far the largest gain for any cohort in any year, see Table 3). However, this phenomenon may also reflect an age effect, or cohort effect, with other grade levels improving less between the two years. In any event, we intended to select as anomalous cases those students whose new scores did not conform to the expected ranking of scores based on the mean NS ranks of his/her cohort (whether or not these progressed from the expected chronological sequence of development). We realized, however, that the very pattern of NS mean ranks for each cohort may reflect the phenomenon of interest; i.e. topic effect influencing task construction, and thereby influencing holistic score as a measure of performance. Thus, we selected case studies solely on the basis of AR anomalies; i.e. AR's which did not progress in a direct line from lowest to highest over the span of four topic/years.

As a final criterion for selection of case studies, we eliminated those cases which seemed problematical, or of less interest for a variety of reasons; e.g. poems, lowest level writers whose papers evidenced scribal difficulties or too little writing to task analyze, highest level writers whose score ranks were skewed due to the "ceiling effect" of consistently high level performance (though some of the latter were included in our sample, if it was possible to make clear forced choice rank decisions about the papers).

### The Typology

Initially, we tried to adapt existing classification schemes to the papers in our sample, and roughly categorized them in terms of "narratives" (stories), "descriptive writing" and "exposition" (commentary, argument). However, we found that this rough classification did not do justice to certain distinctions in the writing sample that were critical for description of task construction in all its complexity. We realized (and other text design research suggests) that students generally seem to have some notion of literary "genres" on which they model their writing, but that they also often combine elements of traditional genre categories to produce a text. For example, we noted that some writers chose to use a narrative "schema" (telling a story) in order to dramatize an idea. Other writers were able to express the same or similar idea, equally successfully, by using expository prose (i.e. a hierarchical arrangement of arguments in an explicit discussion). Similarly, some writers chose to reveal a character "they would like to be" by narrating a key event in that character's life. Others chose to describe the character in terms of its habitual actions.

In order to capture these distinctions, we found it both necessary and useful to separate task construction into two elements, expressed in our typology as "rhetorical function" and "discourse strategy." Essentially, we made a distinction between the form and the function of a text.

By rhetorical function, we mean "What does the text do?" "How does it function for the reader?" Function thus involves both the topic focus and the rhetorical effect of a given text. By discourse strategy, we mean "How is the text structured? What are the means through which a rhetorical function is achieved? What is the form of the text?" Although, as we shall see, certain strategies are conventionally associated with particular functions (giving rise to more traditional text classification schemes), writers can and do successfully employ a variety of different strategies to fulfill particular functions. See Figure 2 and the discussion following.

## Function

There are three analytically distinct functions, or "things that a text can do" (though, see discussion below of phatic and performative text functions not included in our typology): a text may (I) "tell a story," (II) "reveal an entity," and/or (III) "express an idea." Each of these functions is associated with a particular topic focus and rhetorical effect, as mentioned above. In Function I, "telling a story," the topic focus is on a single event or sequence of events, and the rhetorical effect is to relate "change over time." In Function II, "revealing an entity," the topic focus is on a single, concrete entity (object, place) or identity, (character, animal or human, real or fantasy) and the rhetorical effect is to relate "that which remains the same over time." In Function III, "expressing an idea," the topic focus is on an idea, an abstraction, or a classification scheme, and the rhetorical effect is to relate "relationships between entities, events, writer, reader, and/or ideas existing beyond time." Each of these functions is achieved through the use of discourse strategies.

## Strategy

Strategies provide structure for the discourse unit known as the "text." As there are three analytically distinct functions, so there are three strategy types: (A) narration, (B) description, (C) evaluation (commentary). While strategies, as a whole, provide structure for individual texts, the strategies themselves contain an internal structure of elements, termed "propositions," concerning events, descriptions and commentaries. Strategies are defined by the organization of the propositions they contain. Individual propositions in a given text may be identified in terms of formal characteristics. **Event** propositions have the form: "x happened." **Descriptive** propositions have the form: "x is, was, or happens." **Commentary** propositions have the form: "y believes, thinks, feels that x happens, happened, is, or was." Furthermore, commentary propositions may take the form: "z believes, thinks, feels that y believes, thinks, feels that x happens, happened, is or was." That is, commentary propositions may be recursive.

**Proposition Structure.** Although the classification of propositions is sometimes text dependent (e.g. "John is my friend" could be either description or comment depending on whether "John" or "friendship" is the topic focus of the text), there do seem to be certain characteristic markers of each proposition type: Event propositions are marked by use of the **past tense** (except for copulas). Descriptive propositions are marked by use of the **present tense** (and/or **past tense copulas**). Commentary propositions are marked by the following:

1. Explicit statement of opinion, wish, desire, etc. (subjunctive mood?) such as "I wish," "I think," "I hope."
2. Negation, as in "He didn't go."
3. Modals, conditionals (should, would, could, can, be able).
4. Comparatives, superlatives.
5. Abstract subjects (marked by indefinite article or mass noun).
6. Cliches, aphorisms.

It is important to point out here that each strategy--narration, description, and evaluation--contains all three types of propositions. It is not the presence or absence (or relative amount) of proposition types that determines overall strategy but rather the pattern of organization of the propositions, which structures the text, and therefore defines strategy. For example, strategy A (narration) is composed of a chronological linking of "events" which form the spine of the text and through which the text is developed. Events thus serve as the "core propositions" of narrative strategy. In addition, a narration may contain a great many incidental or supportive commentary and descriptive propositions, but these are linked to individual core propositions and not to one another. Similarly, in strategy B (description), the core propositions are descriptive; commentary and event propositions provide support for individual descriptive propositions. In strategy C (evaluation), the core propositions are commentaries; event and descriptive propositions are supportive. See Figure 3 following.

It is also important to note that the organization (linking) of core propositions within each strategy may take several different forms. For example, the event propositions within a narrative strategy may be linked to an overarching event ("the story") and to one another as elements of the "plotted" narration; i.e. protagonist, goal, complication, climax and resolution. Or, event propositions may be linked only to one another, as discrete event units through which a main character proceeds. Each of these event units may contain plotted story elements; e.g. in a "picaresque" narration, or these discrete event units may be linked to one another without containing plot elements, but rather to serve as point of reference, stages, or periods in the "life" of a main character (such as epic, historical, or certain forms of biographical narration).

Similarly, descriptive propositions may be ordered chronologically to produce a sort of "recipe" (a temporally ordered set of procedures), or they may be ordered spatially, associationally, or focally. Commentary propositions may be linked hierarchically (in what we tend to think of as "formal" or structured argument), or they may be linked focally or associationally (see Pringle and Freedman, 1981).

**Primary Types.** As mentioned above, there are conventionalized (standard) associations of particular strategies with particular functions, which we have termed "primary types": Function I (tell a story) with strategy A (narration); Function II (reveal an entity) with strategy B (description); Function III (express an idea) with strategy C (commentary). The "primary type" matching of function and strategy may in fact be expressed in several different ways, depending on the sub-type of strategy used (i.e. depending on the way in which core propositions are linked). For example, a writer may tell a story using several different types of narrative strategy, or may reveal an entity using various styles of description. The following discussion of sample papers will reveal both the basic association of strategy and function in these "primary types," and the range of possibilities within each strategy classification.

Primary type IA (in which a story is told through narration) may be realized through what we have termed the "picaresque" or the "plotted" strategy. Use of a plotted narrative strategy to tell a story may be clearly seen in case #110/73. Picaresque narration is used in case #107/75, on the other hand, to serve the same function of telling a story.

Primary type IIB (in which description is used to reveal an entity) is illustrated in case #420/78 (spatial description), 330/76 (chronological description), 220/74 (focal description), and 326/75 and 108/75 in which different forms of associational description are used

(#326/75 constituting a "self-declaratory" monolog through which the entity is revealed, and #108/75 constituting a "slice of life" depiction of a character).

Primary type IIIC (in which an idea is expressed through evaluation and/or commentary) may similarly be realized in a variety of ways. Perhaps the clearest (or most familiar?) examples of such "expository" writing are those in which the core propositions (comments and evaluations) are linked hierarchically in the form of a logically ordered argument as in case #215/76. It is also possible to evaluate an idea, or comment upon a relationship between things, people and/or ideas, by organizing commentary propositions in an associational or focal way. Often, the combination of associational or focal C strategies with III purpose takes the form of "musings" or "reveries" about a relatively abstract topic as in cases #230/77 and #127/73.

In addition to these primary type combinations of function and strategy, there are a variety of ways in which different strategies may be successfully used to fulfill the same function. As indicated above, it is equally possible to reveal an entity (a character) through a dramatization of that character's life (or a key episode in that character's life) as it is to straightforwardly describe that character. Similarly, one may express an idea by "showing" that idea (through narration) or by "telling" that idea (through commentary, cf. Kaplan and Keech, 1980).

However, one combination of strategy and function that is not theoretically possible, is the use of any strategy other than narration in order to tell a story. That is, Function I can only be fulfilled through strategy A, since a "story" necessarily involves depiction of a chronological sequence of events mirroring change over time in the real world. This seems to be related to a more general pattern of uni-directional constraints on possible combinations of function and strategy. That is, there is a progression in level of abstractness from Function I to Function III; from the experientially-based "revealing an entity," to the abstractness (and lack of temporal reference) involved in "expressing an idea." Similarly, there is a progression from the preponderantly experientially based language of A strategy (narration) through the concrete language of B strategy (description) to the abstract language of C strategy (commentary, evaluation). We noted that the common pattern of combining these aspects is that relatively concrete strategies may be used to fulfill more abstract functions, but that the reverse does not tend to occur. That is, A strategy may be used to fulfill Functions I, II, and III; but C strategy may not be used to fulfill any function other than III (at an equally "abstract" level). Similarly, Function I can only be fulfilled through use of the experientially-based language of strategy A. However, Function II can be fulfilled through use of A or B strategy. This directional

pattern of strategy/function combination can best be understood with reference to the following chart:

I	A	Generalized Experience
II	B	Concrete "facts" from Experience
III	C	Abstraction from Experience

Several examples of strategy/function combinations seemed particularly salient to us, since we have recognized them in literary genres as well as in the student papers under analysis. The combination of IIA (entity and narration), for example, occurs in the form of epic narrations (e.g. of the lives of historical characters) as well as picaresque narrations serving primarily as character sketches as in cases #110/73 and #110/74.

Other examples of combined function/strategy include "mixes" of functions or "dual functions"; e.g. some texts may function both to tell a story AND reveal an entity (Function I-II). This may be accomplished through use of a single strategy (A, for example in case 104/73). Similarly, functions I and III are frequently combined in stories which serve to express an abstract concept or an idea, through narrative strategy (I-II/A combination, as in case 109/74). Finally, dual functions may combine with dual strategies in various ways. Strategies can be mixed through framing (e.g. an evaluative introduction and coda surrounding a narrative body), shifting (a sharp transition from one strategy to another within the body of the text), merging (a subtle transition, back and forth, between two strategies within the body of a text), and integration (where core propositions of two strategies are linked to one another, so that the text is structured in terms of the simultaneous development of two strategies, e.g. narrative and evaluation). Particularly clear examples of dual function/dual strategy combinations can be found in the "autobiographical essay" genre, as in student papers which present the development of an idea (specifically of "self-awareness") over time (I-III/AC combination, see case 302/78). Similarly, description and commentary are often integrated as strategies to fulfill a mixed II-III function (expressing an idea while revealing an entity), though this particular combination usually seemed to appear in student papers as a response to particular topic constraints and as an effect on function of the use of a great deal of abstract language in an otherwise descriptive text serving to reveal an entity.

#### Application of the Typology

The first step in application of the typology to case study papers involved the training of raters and

administration of a reliability check among three raters to ensure the replicability and consistency of the coding process. Raters were trained, as for holistic scoring sessions, through the selection, coding, and discussion of anchor papers chosen as representative "types." Once rater agreement was achieved in these preliminary discussions, a sub-sample of 24 papers was chosen randomly from among the case-study papers: 4 papers, one from each topic/year of each of the six cohorts. Three raters coded papers independently, in terms of function and strategy, assigning codes from 1 to 6 (which encompassed dual functions and dual strategies) to each paper, one number code for function and one for strategy.

We realized, during discussion of the anchor papers prior to the coding session, that raters seemed to agree fairly closely about the most "salient" or "primary" function(s) of a given text, although raters occasionally perceived different secondary or incidental functions within the same text. Similarly, raters were able to agree on salient or primary strategy more easily than on secondary or incidental strategies which they perceived in a given text. Therefore, we decided to compare the independent rater codings in the reliability check in terms of a minimal criterion of agreement: i.e., if raters agreed on at least one element of strategy and one element of function within a given text, we considered that to constitute rater agreement. For example, if rater #1 assigned a function code of III to a text, while rater #2 assigned a code of II-III to the same text, they were considered to be in agreement concerning the primary or most salient function of that text. If two or more raters failed to agree on all aspects of function and/or strategy, we decided that a rater discrepancy existed, and looked to see if this discrepancy could be resolved through a different rater comparison. For example, if rater #1 assigned a function code of II, and rater #2 assigned a code of III, this constituted a rater discrepancy; there was no aspect of function on which they could agree. Out of the 24 papers in our reliability check sample, there were 7 cases in which such rater discrepancy occurred. In every one of these cases, however, discrepancies occurred only between 2 of the raters. Since there were 3 independent raters participating, this also meant that at least 2 out of 3 raters agreed in every single case. Thus, discrepancies between 2 raters occurred in 30 percent of the cases; discrepancies between all 3 raters never occurred (i.e. there were no cases in which no agreement could be reached on both function and strategy); and agreement between at least 2 out of 3 raters occurred in 100 percent of the cases.

Given our experience with the reliability check (and our feeling that the typology was--and still is--subject to some refinement), we decided to code our larger sample of papers according to the following procedure:

1. Each paper was coded independently by 2 raters, who assigned independent codes to function and strategy.
2. Agreement between the two raters was assessed according to the same criteria as in the reliability check; i.e. identical coding of one or more aspects of function and/or strategy constituted agreement on the primary characteristics of that function and/or strategy. Furthermore, if raters agreed on one aspect of, say, function, but disagreed on one or more other aspects, we assigned the agreed upon code to a "primary function category" on our final coding sheets, and assigned the discrepant rater codes to a "secondary function category." For example: rater #1 reports a code of I-III, rater #2 reports a code of I, the primary function code would be I, and the secondary function code would be III; or, rater #1 reports a code of I-III, rater #2 reports a code of I-II, the primary function code is I, and the secondary function code is II-III. In this way, we attempted to preserve a distinction between the most strongly perceived, or most salient function (or strategy) of the text, and those which seemed incidental, ambiguous, less clearly defined, or which may have been the product of rater "error." (Note: The results reported in the following section are based exclusively on primary function and primary strategy codes, for two reasons: so that the effect of rater error may be minimized (through the exclusion of secondary categories, to which discrepant rater codes were assigned), and in order to maximize the possibility of discovering distributional patterns in the data.
3. Cases of discrepancy between the first two raters were assigned to a third independent rater for coding. As in the reliability check, the use of a third rater resolved most discrepancies in the coding process. Including the cases used in our reliability check, our sample totaled 252 papers. Out of that sample, again including reliability check papers, 60 discrepancies occurred in function and/or strategy coding. (If a single paper contained discrepant codes for function and strategy, it was only counted as one discrepant case.) Of the 60 discrepancies, 49 were clearly resolved by the third rater, and 11 were ultimately coded as "unclassifiable" (see #4 below).
4. Finally, where two raters agreed that papers were unclassifiable according to our typology (see discussion below), these papers were coded as "9" (unclassifiable). Similarly, if all three raters failed to reach agreement, or if the third reader agreed with one of the original readers that a paper was unclassifiable, or if the third reader alone felt the paper was unclassifiable (and that this, therefore, was the source of original rater discrepancies), the paper was coded as a

"9." Since function and strategy were coded independently of one another, it should be noted that a "9" function code did not necessarily entail a "9" strategy code (again, see below for discussion of these cases).

### Problems in Application of the Typology

The most frequent source of discrepancy in coding appears in papers which, according to raters, seemed to contain some combination of description and evaluation in function and/or strategy (22 out of 60 discrepancies in the total sample). For example, there were papers in the sample which seemed to contain a great deal of descriptive language (B strategy) along with an abstract topic focus (generally indicating a III function). Raters differed in coding these papers as II/B, III/C, II/B, or even II-III/B, C, or BC. Where a discrepancy existed, the third rater generally provided a II-III function code, which resolved the discrepancy. When one rater indicated a II-III function, where the other indicated III function, where was no discrepancy and the paper appears in the final coding as having a III primary function and a II secondary function (as opposed to the II-III primary function provided through third reader resolution). Thus, several essentially similar papers appear in our final coding as either II-III or III function papers (having either a B or a BC primary strategy). Similarly, several papers seemed to contain relatively abstract, evaluative language (C strategy) along with a concrete, descriptive, "character focused" function (II). Again, these papers appear in the final coding generally as II-III/BC types. However, some were finally coded as II/C's and some as II/P's with III and C as secondary function and strategy.

These discrepancies, and the general ambiguity in the final coding of such papers, seem attributable to two factors: topic effect, and what we have termed the "bleeding strategy effect" on rater perceptions. Topic effect influenced the coding process (as distinct from the distribution of task constructions, discussed below in "Results"), to the extent that some writers appeared to be attempting to honor topic constraints "independently" of the primary function and strategy they established in a given task. This was especially clear in responses to topic/years 1974 and 1975, where students were asked to write about a "non-human thing" (1974) or a "person other than oneself" (1975) that they "would like to be." Students appeared to vary in constructing the task so that either: 1) most emphasis was placed on "describing" the character, with a nod to topic constraints consisting of some evaluation of the desirability of being that character; or 2) emphasis was placed on either describing or evaluating the writer's "feelings about being that character," with a nod to topic constraints in the form of descriptive statements about who or what that character was. Depending on the relative emphasis a writer placed on the

central character or on the writer's feelings about being that character, and depending on whether the writer emphasized description or evaluation of his abstract or concrete subject, raters seemed to vary in assigning codes to these papers; i.e. all these papers seemed to contain elements of description and evaluation, which different raters perceived as having different degrees of importance to the overall text. Ideally, perhaps, all papers of this sort should have been coded as II-III/BC papers, in recognition of the combination of elements they contain. However, this "type" of task construction was not so clearly identified as some others in discussions prior to the coding process, so some ambiguities remain in the final codes.

A related topic effect influence on coding can be found in papers coded as II/B's with C secondary strategies. These papers (again, generally written in response to topic/years 1974 and 1975) were essentially "character sketches," or descriptive language used to reveal an entity. However, because of the "hypothetical" wording of the topic, many writers used relatively abstract, "hypothetical" language in their descriptions; i.e. a great many modals, such as "would" and "could," as in: "If I were an eagle I would build a nest high in the mountains." This use of modals influenced some raters to assign a C strategy code to these essentially descriptive papers, and, in fact, the explicit honoring of topic constraints which dictate the use of modals does distinguish these papers from others in which "purely" descriptive language was used (e.g., "I am an eagle and I always build my nests high in the mountains"). However, in the report of results given below, these papers having a C secondary strategy will not be distinguished from "pure" II/B types, since only primary category codes were analyzed.

The "bleeding strategy" effect on rater perceptions (and therefore on the coding process) is related, we believe, to the uni-directional pattern of function/strategy combination outlined above. That is, there appears to be a constraint operating on function/strategy combination such that relatively abstract functions may combine with relatively concrete strategies, but the reverse is not true--concrete functions cannot be fulfilled through the use of abstract strategies. Thus, a story cannot be told solely through description or commentary, but must be told through narrative. Similarly, commentary, description and narration may each be used to express an idea, but commentary (ideally) cannot be used to reveal an entity. Therefore, it seems that when a writer does use a great deal of evaluative language (many comparisons, or the use of abstract nouns, etc.) to fulfill a function which is primarily one of "revealing an entity," the rater is generally inclined to perceive a more abstract function in the text (i.e. that it "expresses an idea"). This is not to deny that function and strategy are empirically separable--the fact that they are

is evidenced in several sample papers which raters agreed should be coded as II/C rather than II/B or III/C. However, many of these papers (which evidenced the "bleeding strategy" effect) were ultimately coded as II/BC, III/BC, or II-III/B, C, or BC. Again, essentially similar papers (similar in terms of task construction) were finally coded in a variety of ways, though it is important to point out that the essential distinction between II/B and III/C task constructions was maintained in the final codes. That is, none of these "ambiguous" papers was ever coded as a III/C, although the III/C coding category does contain an internal range of variation in types which is quite distinct from that just discussed.

Essentially, the III/C code was applied only to papers which expressed an idea through the use of commentary or evaluative language. However, papers in this general category varied considerably in terms of the arrangement of commentary propositions they contained. Some of these papers consisted solely of associationally linked "musings" about a particular idea (e.g. "the danger of guns"). These might be termed "stream of consciousness" essays, in which the author appears to be sustaining one end of an open-ended conversation, rather than presenting a clear sequence of logically (hierarchically) linked arguments. In this sense, they constitute what Keech has termed "engaged dyadic (open schema) strategies" (cf. "Toward a Developmental Theory of Discourse Aims and Strategies," in progress). Other, similar, papers in the III/C category were organized focally around a specific topic or idea; for example, consisting almost entirely of a series of rhetorical questions concerning that idea, addressed to a fictive audience including and extending beyond the "test reader." For the purposes of this analysis, however, it seemed most important to stress the commonality of these papers in coding them; they all seem to represent stages in the development towards "classic" expository prose, or a hierarchically ordered sequence of arguments through which the author anticipates and responds to a range of possible interlocutors (the readers of his text). Furthermore, it is important to point out that the use of an associationally linked strategy does not preclude the production of expert, highly proficient and successful prose. There are degrees of proficiency within the "stream of consciousness" style, as there are in "classic exposition or argument," and in fact within each type of task construction we have presented here. Thus, the grouping of different commentary strategies under the III/C category does not obscure differences in quality levels within that category any more than does the grouping of variously "good" and "bad" papers under any other category of task construction.

Aside from the above mentioned distinctions, and similarities, among papers that may have been obscured in the coding process (or that gave rise to discrepancies in coding), there was another relatively large group of papers

for which our typology could not adequately account. These papers were coded as "unclassifiable," in terms of function or strategy or both. Papers were unclassifiable for one of two reasons: either they appeared to fulfill more than two functions and/or strategies simultaneously (and therefore contained no clearly defined primary function or strategy), or they fulfilled a social or phatic function (through use of a distinct strategy not included in our typology which overrode any rhetorical function included in our typology).

Examples of the first type of unclassifiable paper include those finally coded as II/9's; i.e., those with a II function, but an unclassifiable strategy. These papers were generally written in response to topic/year 1977 ("choose an object to take into space"). Typically, these papers focused on the decision-making process (deciding which object to take) rather than on the qualities of the object itself, or an evaluation of that object as a "good choice." We classified these papers as having II functions because the topic focus seemed to be on the main character (the "self") who was (or had been) making the decision. The strategy, such as it was, seemed to be to mirror the thoughts going through the mind of that "self" during the process of decision making (selecting from among possibilities, rejecting and/or accepting a range of possible choices). In most cases, the language of this strategy is past tense, and appeared to be "narration"; yet, it was also very evaluative and descriptive of the objects being considered. These papers were actually quite interesting examples of what may be a distinct task construction type. They appear, superficially, to be similar to the autobiographical essays, coded as I-III/AC. However, where the autobiographical essay traces the development of an idea over time (culminating in an increased self-awareness or understanding), the II/9 paper simply mirrors a thought process--the self as topic focus is a static entity, the "growth" or change is a process of rationalization resulting, not in a change in personal attitude, but merely in a decision made. It is important to point out that these writers are not justifying a decision that has been made (as in a II/C paper, or a II-III/BC in which the object is fully described as well as evaluated); rather, they are reporting the process of making a decision. Narrative past tense and narrative present tense papers are alike: "First I thought--then I thought" compared with "Well, I could take ----. But it needs to be -----."

Examples of the latter groups of unclassifiable papers include letters, speeches, and certain "dialogues." These papers were coded as "9's" in terms of function and strategy. As mentioned, the function of such papers seems primarily to be one of establishing social rapport, communion, or solidarity. They are addressed to specific "respondents," and usually contain content of little or no significance, except as a focus for the expression of

concern, affection, interest, or gratitude on the part of the writer towards another person or persons. For this reason, we have deliberately excluded these papers from classification according to our typology of "rhetorical discourse functions and strategies." Rather, these papers seem to primarily fulfill either "phatic" functions (e.g. letters to loved ones, friends or acquaintances; speeches to a specified audience rather than the "test-reader") or they fulfill "performative" functions (e.g. thank-you letters, job applications)--all of which are oriented more towards establishment of a social relationship between writer and fictive audience, than towards persuasively communicating information about facts of the world.

### Results and Hypotheses

As a first step in this research, we looked for correlations between topic and score in terms of the pivotal variable of "word count." We discovered that, as had been predicted, word count correlates significantly with new score (NS) for our sample ( $n^2 = .10$ ). New score tends to increase with word count. Therefore, we hypothesized that word count might be increasing with topic (as a correlate of task construction used within each topic), thus providing a pivotal link between topic and score. However, we discovered that although a trend exists for higher word count to equal higher score, the variance of word count within each score was too great to cite word count as a major factor. Furthermore, word count did not seem to be affected by topics in consistent ways; e.g. topic/year 1973 generated the fewest words, but the second highest mean score, while topic/year 1976 generated the most words, but the fourth highest mean score out of six topic/years).

We next looked for correlations between topic, grade and task construction as a preliminary to discovering possible relationships between topic and score. Since we ultimately wish to examine the factor of task construction as a variable influencing quality ratings (or change in an individual student's quality ratings over time), we felt it was first necessary to determine which task constructions were favored for a given grade within a given topic/year. Then we will be able to tell through further analysis whether a student's use of an unexpected (or "unfavored," "unpreferred") task construction in a given year might have affected that student's score (i.e. was that student trying to do something new or different both from what he had done previously, and from what other students generally did in response to the same topic? Did this influence his score?) Future analysis, then, will concentrate on case studies of students who deviated from the "favored" task construction in a given topic/year, as compared with students who did produce the favored type.

The results as reported here will make reference to the

variables of topic/year, grade (freshman through senior) and task construction. For purposes of computer analysis, task construction was coded in terms of independent variables of function and strategy (f/s). These are reported in the following discussion in terms of two-digit codes: 1,1 = I/A; 2,2 = II/B; 3,3 = III/C; 4,4 = I-III/AC; 5,5 = II-III/BC; 6,6 = I-II/AB (plus various permutations of these codes, such as 3,4 = III/AC; 5,3 = II-III/C; etc.). The code for unclassifiable function or strategy is "9," as discussed above.

Topic/Year 1973. "Event" Topic. In the initial examination of the 1973 prompt, and early examination of student responses, it was noted that there were two distinct ways students seemed to interpret the instruction: "Write about an event you wish you had witnessed or wish you could witness." Some students seemed to understand the command "write about" to mean identify (such an event) and comment on it. Others seemed to understand the assignments to allow relating or recounting the event, possibly narrating it as if it had actually happened, or even as if it were happening as the narrator spoke. It would appear that a writer had made the first choice if he produced an informal essay using strategy C (commentary) and having the function of expressing an idea (III); in this case, investigating the writer's reasons for wishing to witness the event. The effect of the second choice would be to produce a story (I) using strategy A (narrative). I/A and III/C, then, might be identified as predicted types of response for this topic. The results in fact show that these are the two preferred types of response.

In 1973 overall, 43 percent of the responses were type I/A (1,1) while 34 percent were III/C (3,3). All other responses (totalling 23 percent) were distributed over nine different categories, no one of which contained even 5 percent of the 1973 student responses. These unpredicted, or less preferred types of writing are of passing interest, however, since six of the nine alternative types included a narrative strategy (e.g. 4,1; 2,1; 9,1; 6,1; 3,1; 2,4). Although numbers in each cell are small, a clear trend is evident if preferred strategies are compared from freshman to senior year. Fewer freshmen use type I/A (1,1); fully half of them preferred type III/C (3,3). Sophomores are about equally divided in their preference for the two strategies (30 percent choosing 1,1; 40 percent choosing 3,3), while juniors shifted even further towards a preference for type 1,1 (50 percent), leaving 30 percent choosing 3,3. Finally, seniors overwhelmingly preferred 1,1 responses to this topic, seven out of ten (70 percent) choosing to tell a story, leaving only three students choosing other solutions and only one of them (10 percent) writing a 3,3 type paper. On no other topic did as many as 20 percent of the seniors choose telling a story (1,1) as their mode of response; out of a total of eleven seniors writing 1,1 type papers on all

topics combined, 7 of them (64 percent) were written in response to the 1973 (Event) topic.

#### Topic/Year 1974. "Be a non-human thing" Topic

The 1974 topic produces fewer dominant preferences. Only among freshmen is there a strong central preference: type II/B (2,2) or descriptive strategy used to reveal an entity. 50 percent of the freshman papers on this topic were type 2,2 while of the remaining 50 percent of the papers, 32 percent contained some II function. There is also a weak preference among sophomores for the 2,2 type (30 percent). Juniors and seniors in this topic/year, however, evidenced no clear preference for this type; seniors in fact produced as many 3,3 papers as they did 2,3's (both at 20 percent, and both greater than the percentage of 2,2 papers, at 10 percent). Interestingly, this topic/year was the one in which mean scores within each cohort were highest; i.e., students in all grades writing on this topic, on the average, tended to get their highest scores here.

#### Topic/Year 1975. "Be another person" Topic

We expected, initially, that patterns of task construction distribution for the 1975 topic would be essentially identical to those in 1974, since the topic constraints (expressed in the prompt, or "given task") are so similar for each topic/year. However, 1975 responses show a clear pattern of preference, changing over time (by grade), unlike 1974 responses. In fact 1975 responses demonstrate what we hypothesized as the expected developmental progression, from 1,1 to 3,3 (freshman to senior, experientially-based writing to abstract exposition). Freshmen in 1975 show a weak preference for stories (1,1: 27 percent), which shifts among sophomores to a weak preference for description (2,2: 25 percent). The amount of evaluative responses (3,3) stays relatively constant from freshmen to sophomores (at 18 percent and 16 percent respectively). Among juniors, however, a clear shift towards the 3,3 type occurs (30 percent compared to only ten percent of the 2,2 type within this cell). Seniors writing on this topic complete the pattern of shift towards the 3,3 type, where it occurs in 5 out of 11 papers (46 percent, along with C strategy in 20 percent, and BC strategy in 20 percent of the remaining papers in the cell). One hypothesized explanation for the stronger trend towards 3,3 task constructions in 1975 as compared to 1974, is that the topic constraint of writing about a person, rather than a non-human "thing," favored the use of relatively abstract language (since a person necessarily has thoughts and feelings which must be described somewhat abstractly). Put another way, it may be that the requirement of writing about a "thing" favors the use of relatively concrete, descriptive language, and relatively more emphasis on the "thing" itself than on one's reasons for wanting to be "it."

### Topic/Year 1976. "Invention" Topic

We originally hypothesized that topic/year 1976 would strongly favor the use of a 3,3 type task construction, since the given task explicitly calls upon the writer to evaluate an object, and support that evaluation through some form of evaluation. Indeed, 3,3 was overwhelmingly preferred by writers within each grade, though preference increased from 50 percent among freshmen to 80 percent among seniors. Interestingly, too, the range of alternatives chosen by writers decreased notably from freshmen to seniors. Where freshmen responses varied among 5 alternatives to the 3,3 preferred type (making up 10 percent of each cell total), sophomore and junior responses were limited to 4 alternatives, and senior responses included only two alternatives (one story and one II/BC). It is worth noting here that those writers who did choose some combination of 2 function with a 2 or 3 strategy, did so at the potential cost of ignoring a strong topic constraint. These writers focused almost exclusively on the "invention" itself (e.g. detailing its history) in violation of the constraint: "Make it clear why humanity would be better off without this invention."

### Topic/year 1977. "Space object" Topic.

Again, we expected to find a preference for 3,3 in this topic/year, since the given task calls on the writer to "recommend" an object and support that recommendation. As expected, 3,3's are strongly favored in this topic/year, though not so strongly as in topic/year 1976. Freshmen show an equally strong preference for 3,3 in 1977 as compared to 1976 (50 percent), but sophomores show a weaker preference (40 percent in 1977), and their responses include 6 alternative types as opposed to 4 alternatives in 1976. Juniors show a much weaker preference for 3,3 in 1977 (27 percent) and also show a weak preference for 2,9's (27 percent, discussed above as the "thought mirror" type). Seniors again show the strongest preference for 3,3 of any of the grade levels in response to this topic.

One aspect of the different pattern of response to 1977 as compared to 1976 topic/years, is that many writers seem to have found a "diary" format appropriate to this task, and their responses include several who chose to discuss "how they came to be chosen for the space flight" or "what the experience of space flight it like," as well as those who discussed "how they came to make a decision about the object." Thus, there were somewhat more stories written overall for this topic as compared to 1976 (12 percent vs. 9 percent) and a greater range of variation (including 9,9 types, and various combinations of narrative strategy with II (2) or III (3) functions).

## Topic/Year 1978. "Influence" Topic

This topic generally seemed to be the most complex in terms of given task constraints. We therefore hypothesized that it would elicit the widest range of task constructions of any of the topics; e.g. 1) if the writer responded by focusing on "the influence" (effect on life, changes in life), the text would probably have a III function, 2) if the writer focused on the agent of influence, he or she might produce a text with function II, although 3) if the agent discussed is an event, a text with function I might be produced, and 4) if the writer focuses on both the agent of influence (an event) and on the influence itself (e.g. change in internal state of the author, change in self-awareness), the text produced would be an "autobiographical essay," having a I-III function. In fact, responses to topic/year 1978 do show the widest range of variation of any topic/year. The only strong preference shown ins for 3,3 among sophomores (50 percent). Freshmen show a weak preference for 3,3 as well (30 percent); juniors show no clear preference at all (though 3,3 and 4,4--the autobiographical essay--are both represented at 20 percent). Seniors show weak preferences for both 3,4 and 9,9 (i.e. totally unclassifiable papers) at 27 percent each. Overall, stories are represented by 15 percent of the papers, autobiographical essays by 10 percent (greater than for any other topic/year), entity descriptions (2,2) by only 7 percent, and commentary/evaluations/ideas (3,3) by 29 percent.

Finally, to summarize the overall results, the preferred task construction by topic/year is as follows: 1,1 for 1973 (43 percent); 2,2 for 1974 (26 percent); 3,3 for 1975 (27 percent); 3,3 for 1976 (60 percent); 3,3 for 1977 (47 percent); 3,3 for 1978 (29 percent). As we expected, preferences in each case reflect "primary types," or conventionalized associations of function and strategy. Task construction correlate with grade, overall, is heavily influenced by the preference for 3,3 among all grades in 1976, so that 3,3 appears as the preferred type for all grades, at around 33 percent of each. However, type 3,3 does climb to 40 percent of the total senior papers, suggesting that our prediction of development from I to III with age may have some validity.

The 1,1 type, which we expected to be most preferred by freshmen, remains a fairly constant 13 - 18 percent of the total within each grade. This seems due, in part, to the strong preference for the 1,1 type at all grade levels in response to the 1973 topic. It is only in response to the 1975 topic that we can see a clear possibility of grade effect on task construction, which will have to be taken into account in future case study analysis.

For the present, we can say that it is possible to identify more or less strongly preferred task constructions

for each topic/year, and to hypothesize about the interpretation of given task constraints which gave rise to those preferred types. The next step in our analysis will be to determine whether a significant correlation exists between holistic score and the use of the preferred task construction in a given topic/year, within a given grade level. In this way we hope to account for some of the variance in new score not adequately explained by reference to word count. Furthermore, we hope to be able to demonstrate, through case study analysis, that what a student actually does in response to a given topic will influence the quality rating of his work--not because some task constructions are inherently "better" or more "successful" or more "appropriate" than others, but because the individual student may be more successful at using a "tried and true" task construction, than he is when attempting something new and perhaps more complex or challenging.

## REFERENCES

- Bain, A. English composition and rhetoric. American Edition, Revised. New York: D. Appleton & Company, 1890.
- Bereiter, C. & Scardamalia, M. From conversation to composition: the role of instruction in a developmental process. To appear in R. Glaser (Ed.), Advances in instructional psychology. Vol. 2. Hillsdale, N.J.: Erlbaum. (Ms., draft: November 6, 1979.)
- Bever, T.G. Regression in the service of development. To appear in T.G. Bever, et al. Regressions in development: basic mechanisms and phenomena, in press. (Ms., draft: 1981.)
- Breland, H.M., & Gaynor, J.L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, summer 1979, 16 (2), 41-49.
- Britton, J. Language and Learning. Urbana, Illinois: National Council of Teachers of English, 1975.
- D'Angelo, F.J. Modes of discourse. In G. Tate (Ed.), Teaching Composition: 10 bibliographical essays. Fort Worth, Texas: Texas Christian University Press, 1976. 111-135.
- Diederich, P. Measuring growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.
- Freedman, A., & Pringle, I. Writing in the college years: some indices of growth. College Composition and Communication, XXXI: Oct. 1980. 311-324.
- Godschalk, F.I., Swineford, F. & Coffman, W.S. The measurement of writing ability. New York: College Entrance Examination Board, 1966.
- Goodman, D.R. Stage transitions and the developmental trace of constructive operators: an investigation of a neo-Piagetian theory of growth. (Doctoral dissertation, York University, Toronto, Ontario, 1979. Investigates the problem of uneven development applying the theories of Juan Pascual-Leone.)
- Hopper, P. Aspect and foregrounding in discourse. Paper presented at the Symposium on Discourse and Syntax, UCLA, November, 1977. Unpublished ms. available from author, Linguistics Program, State University of New York, Binghamton.
- Keech, C., and McNelly, M.E. Comparison of student and teacher rater responses during holistic scoring: differences in task interpretation. In C. Keech, et al. Effects of variation in a writing test prompt on holistic scores and composing process. (Part Two.) Technical Report No. 2 of the NIE Writing Assessment Project, Marcia Farr, NIE Project Officer, Reading and Language Studies. Grant No. NIE-G-80-0034. Berkeley, California: Bay Area Writing Project, 1982.

(References: continued, p. 2)

- Kinneavy, J. L. A theory of discourse: the aims of discourse. Englewood Cliffs, N.J.: Prentice Hall, 1971.
- Labov, W. & Walletsky, J. Narrative analysis: oral versions of personal experience. In J. Helm (Ed.). Essays on the verbal and visual arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society. Seattle: distributed by the University of Washington Press. 1966.
- Meyer, B.J.F. The organization of prose and its effects on memory. Amsterdam: North-Holland Publishing Company, 1975.
- Miller, S. Rhetorical maturity: definition and development. Reinventing the rhetorical tradition. A Freedman and I. Pringle (Eds.), University of Central Arkansas: L & S Books, 1980, 119-127.
- Moffett, J. Integrity in the teaching of writing. Phi Delta Kappan, December 1979, 276-279.
- Murphy, S., Kinzer, C., & Carroll, K.C. A study of the construction of the meanings of a writing prompt by its authors, the student writers, and the raters. Technical Report No. 4 of the NIE Writing Assessment Project, Marcia Farr, NIE Project Officer, Reading and Language Studies. Grant No. NIE-G-80-0034. Berkeley, California: Bay Area Writing Project, 1982.
- Piaget, J. The development of thought: equilibration of cognitive structures. New York: The Viking Press, 1977.
- Spandel, V. & Stiggins, R. Direct measures of writing ability: issues and applications. A report of the Clearinghouse for Applied Performance Testing. Portland, Oregon: Northwest Regional Educational Laboratory, 1980.
- Stein, N. & Glenn, C.G. An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.). New directions in discourse processing (Vol. 2). Norwood, N. J.: Ablex Publishing Corporation, 1979.
- Williams, J. Defining Complexity. College English, Vol. 40, No. 6: Feb. 1979.

#### Addendum

- Freedman, A. & Pringle, I. Why children can't write arguments. (Linguistics department, Carleton University, Ottawa, Ontario.) Unpublished ms. 1981.
- Moffett, J. Teaching the Universe of Discourse. Boston: Houghton Mifflin, 1968.
- Moffett, J. & McElheny, K. Points of View. New York: New American Library, 1966.

## APPENDICES

### TABLE 1

#### Topics Used in Six Years of Drake Writing Assessment

- 1973 Write about an event you wish you had witnessed or could witness. The event can be real or imagined; the time of the event can be past, present or future. Make it clear why the event is significant to you. You may write a journal entry, letter, dialogue, monologue, essay, story, autobiography, or other form.
- 1974 If you had to choose to be something other than a human being, what plant or animal or other form would you choose? In your writing, give your reader some idea of what you think it would be like to be that form, and of why you chose it. You may do this writing as a journal entry, a letter, a dialogue, a story, an autobiography, an essay, a poem, etc.
- 1975 If you could change places with someone else, who would it be? The person you write about can be living, dead, drawn from past or present, from books, films, etc., or from your own imagination. In your writing give your reader some idea of what it would be like to be that person, and of why that life appeals to you. You may do this writing as a journal entry, character sketch, dialogue, letter, story, autobiographical essay, argument, poem, or other form.
- 1976 Not all inventions have been good for all humanity. Name one invention we would be better off without, and make it clear why. You may do this writing as an essay, journal, letter, story, or other form.
- 1977 Imagine that a small group of people will be sent to colonize a new planet. Food, Clothing, shelter and transportation have been provided for. You are among those asked to select a few additional things to be sent along in the limited space available in the ship. What one item would you recommend, and why? You may write your recommendation in the form of a story, a dialogue, a letter, a speech, an essay or other form.
- 1978 Write about some way in which your life has been, or might be influenced. You might write about the influence of another person, a book or film, an idea, or an event such as a triumph or defeat, or a sudden gain or loss. Make it clear just what or who influenced you, and what the effect was upon you, or what the effect could be upon you. You may do this writing as a journal entry, character sketch, dialogue, letter, story, autobiographical essay, or other form.

TABLE 2

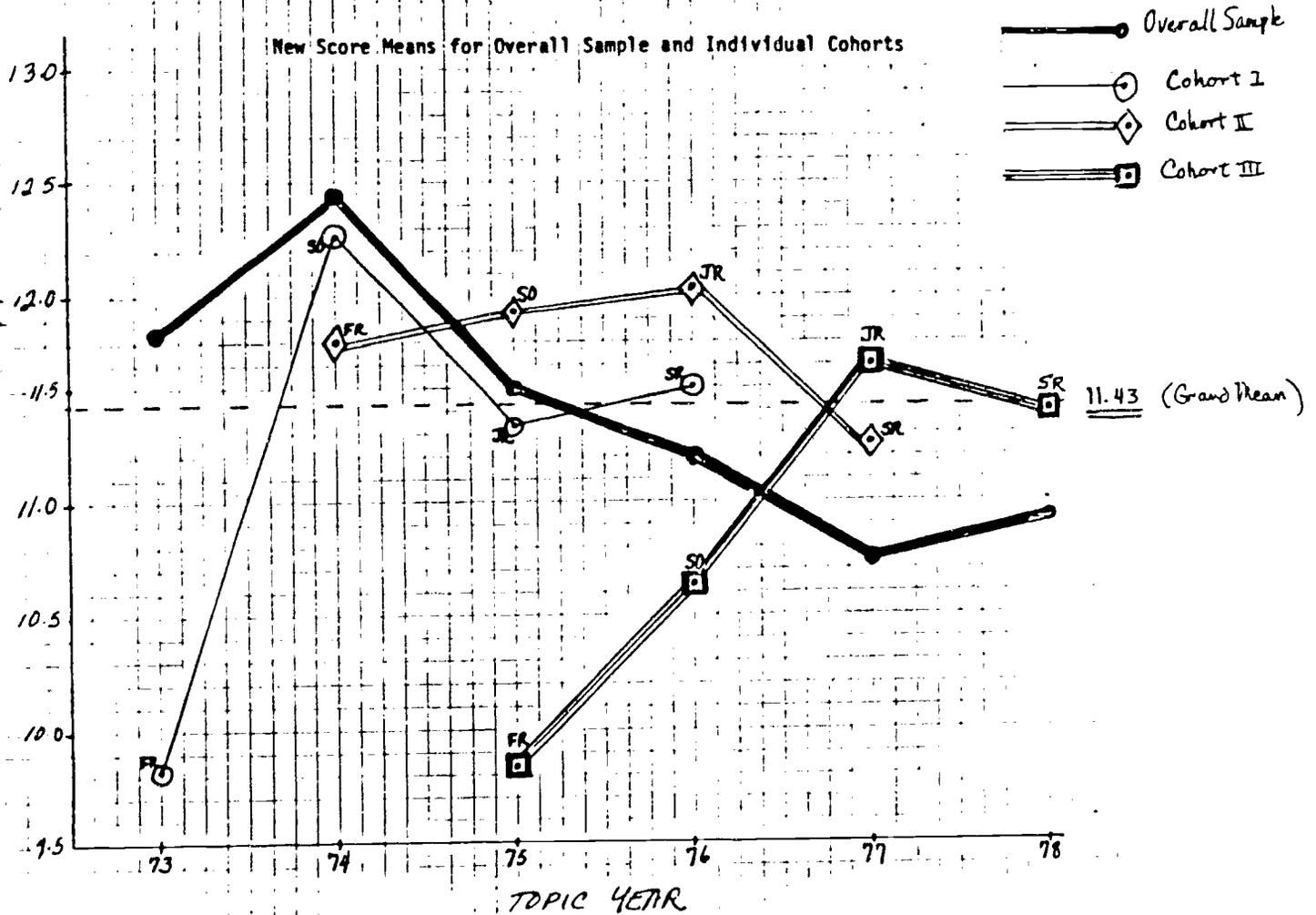


TABLE 3

New Score Means for each Grade Level compared to Over-all Sample Means for each Topic/Year

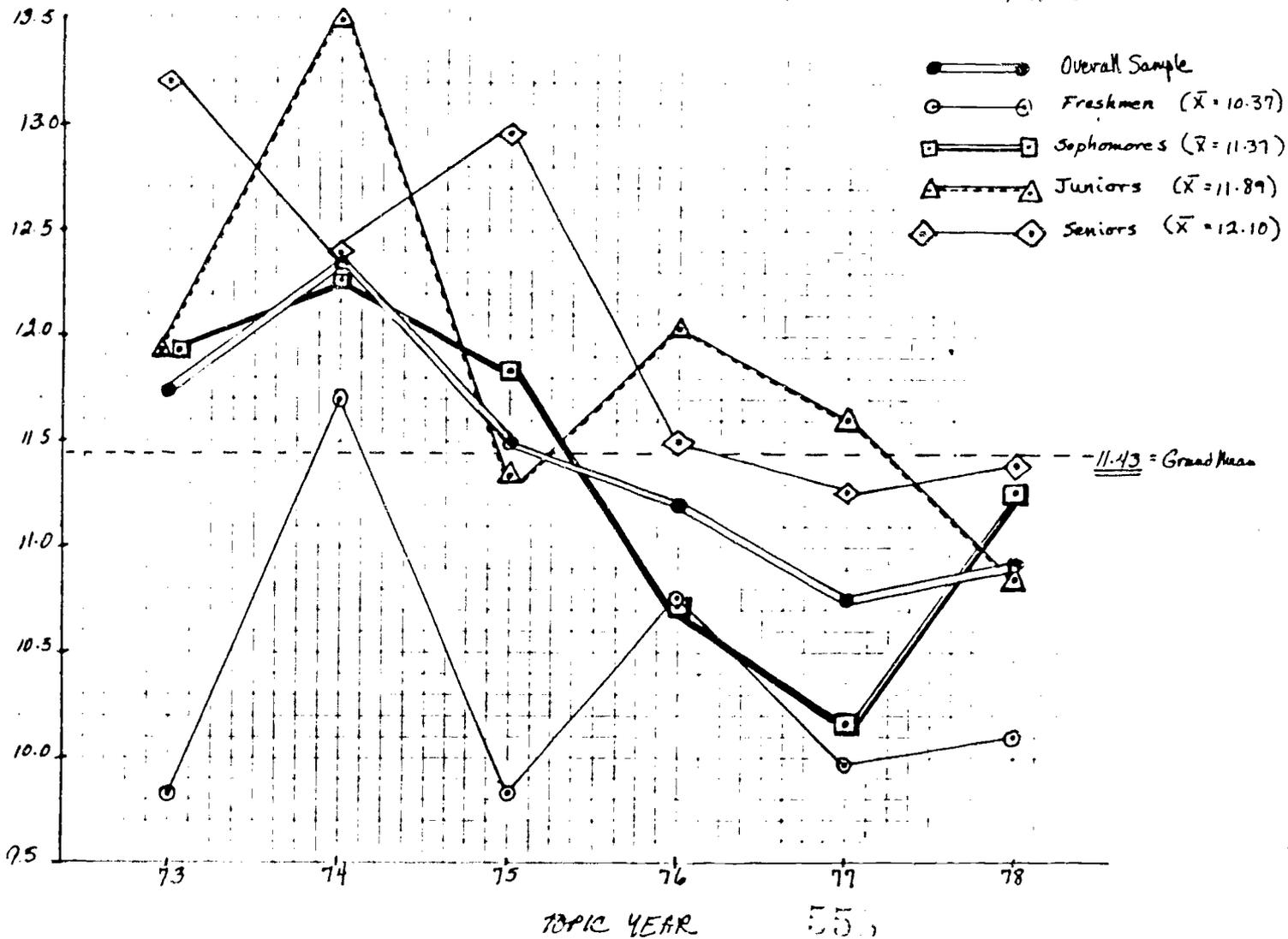


TABLE 4

## Topic Years according to Cohorts

	Grade	9	10	11	12
Cohort I		73	74	75	76
II		74	75	76	77
III		75	76	77	78
IV		78	73	74	75
V		77	78	73	74
VI		76	77	78	79

559