(Unpublished Ph.D. Dissertation)

THE INFLUENCE OF THE TASK UPON WRITING PERFORMANCE

YOUNG MOK PARK

University of Illinois at Urbana-Champaign, 1987

THE INFLUENCE OF THE TASK UPON WRITING PERFORMANCE

BY
YOUNG MOK PARK

B.S., Seoul National University, 1970
M.S., Seoul National University, 1978

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1987

Urbana, Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

NOVEMBER 1986

WE HEREBY RECOMMEND THAT THE THESIS BY

YOUNG MOK PARK

ENTITLED THE INFLUENCE OF THE TASK

UPON WRITING PERFORMANCE

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Barak Rosenshine
Director of Thesis Research

Nancy S. Cole
Head of Department

Committee on Final Examination†

Barak Rosenshine
Chairperson

Robert Linn

Ann C. Peters

Charles Isbell

† Required for doctor's degree but not for master's.

O-517

THE INFLUENCE OF THE TASK UPON WRITING PERFORMANCE

Young Mok Park, Ph. D.
College of Education
University of Illinois at Urbana-Champaign, 1987
Barak V. Rosenshine, Advisor

In an effort to determine whether two different topics (a writing task with general information provided and a writing task with narrowly defined specific information provided) will elicit different qualities and syntactic characteristics of writing from a sample of upper college level students, this study raised three major questions: (a) do the two different topics significantly affect students' writing performance in the syntactic characteristics and quality ratings in a testing situation? (b) are the topic effects consistent across different groups of writers (different nativelanguage groups and different academic-major groups)? and (c) do the relationships among different measures of writing skill remain stable across the two different topics?

For the purposes of this study, ninety six upper-level college students' writing samples (192 essays) and four different kinds of writing-skill measures (elaboration length, syntactic complexity, analytic scores, and holistic scores) were used. To investigate the topic effect, interaction effect, and relationships, two kinds of analysis methodt were used: the repeated measure of analysis of variance method and the correlational analysis method.

The results of this study showed following major points: (a) the general writing task facilitates more elaboration and higher quality in the analytic content; thinking main criteria than the specific writing task, (b) the specific writing task facilitates higher quality in the holistic scoring method, (c) both tasks facilitate almost the same degree.

of syntactic complexity and the same quality in the analytic organization and style/tone criteria, (d) the interaction of the native-language groups and the topics is significant in the elaboration-length variables, (e) the interaction of the major groups and the topics is significant in the holistic scores, (f) the correlation between the essay-quality variables and the syntactic characteristic variables are stable across topics, and (g) the correlations between the analytic scores and the holistic scores vary from topic to topic.

ACKNOWLEDGEMENTS

I am indebted to many people who have not only contributed to the completion of this dissertation, but, more importantly, have influenced my academic and intellectual growth. With sincere thanks and appreciation for the roles they played, I wish to extend my gratitude to the following persons without whom this dissertation could not have been completed:

To professor Barak V. Rosenshine, thesis advisor, for his special help, guidance, and understanding,

To professor Alan C. Purves for encouraging me to investigate the effects of different writing tasks on writing performance and for providing me the data used in this study,

To professor Robert L. Linn for his helpful comments on the research design and method, and for his excellent teaching of advanced statistics,

To professor Charlene Tibbetts for introducing me to research in writing, for hiring me as her graduate assistant, and for supporting my entire work for this dissertation.

I also wish to extend my thanks to Dr. Sybil B. Carlson for providing me the background information on the data used in this study, to Dr. Anna Soter for her helpful comments concerning the proposal, and to Dr. Gail E. Hawisher for sharing her wisdom in English composition with me.

Finally, I wish to thank my family. My mother prayed for the completion of my doctoral study. My daughter, YuJong, and my son, HyongJu, were patient enough to excuse their father's role during writing this dissertation. Without the love and patient endurance of my wife, my doctoral work would have been impossible.

TABLE OF CONTENT

CHAPTER I
INTRODUCTION

General Concern

The difficulty of finding valid assessment measures for students' writing performance has become an important issue over the years. Recent interest in the assessment of writing performance derives from several sources such as public concern over the "writing crisis," lack of a uniform standard for evaluating writing competence, and lack of reliable information on proper assessment procedures. Based on several reports about the decline of student writing ability, the public has demanded accountability.

In order to satisfy the critics and to find a valid assessment of writing performance, several methods have been suggested and reevaluated. Two distinctive assessment methods--direct and indirect--have been used. The direct assessment method requires students to write an essay, and such essays are read and independently scored by two or more readers. The indirect assessment method, on the other hand, requires no writing at all, and the student responds only to items in a multiple choice test. While measurement specialists usually suggest that the indirect (objective) method is more efficient and reliable, most writing researchers suggest that the direct assessment method is more appropriate because the only way to assess whether a student can write well or poorly is to assign him or her a writing task and look at the results.

The indirect assessment method is economical and reliable. Student response on an objective test can be quickly and reliably scored, and the tests have good predictive validity. But the method cannot measure the students' ability to accomplish important

tasks of writing such as the generation and organization of ideas or selection of an appropriate discourse mode. Because of these limitations of the indirect assessment method, most writing researchers agree that direct assessment is the most appropriate and successful means of evaluating student writing performance. Therefore direct writing assessment is currently used for the purpose of providing information about performance of large groups of students and also as a diagnostic measure for improving classroom instruction.

Among the direct assessment methods available, holistic scoring (sometimes called general impression), analytic scoring, and primary trait scoring methods are currently used. The holistic scoring method depends upon an overall or holistic impression of the writing sample, and writing is viewed as a unified, coherent whole. In contrast, in the analytic scoring method, distinctive characteristics or established criteria are emphasized. Each criterion, such as categories of ideas, organization, style, and mechanics is scored separately. The primary-trait scoring method focuses on an analysis of situation-specific traits and provides specific information on a narrowly defined writing task. In addition, syntactic complexity measures, such as T-unit length or the ratio of free modifiers, are also used. T-units (minimal terminable unit: a group of words which constitute a main clause in addition to all subordinate clauses attached to it) and free modifiers (openers, interrupters, and closers) are used for determining syntactic complexity or writing maturity. It is widely accepted that the mean T-unit length increases with maturity and a mature style has a relatively high frequency of free modifiers, especially in the final position.

The direct writing assessment method usually needs several procedures for its success. The assignment must be developed, the scoring criteria selected, the raters trained, the

papers scored, and, fmally, the results analyzed. Writing performance can vary, depending upon the different assignment types (purpose, topic, required discourse mode, audience, etc.). Among these procedures, the assignment development process may have diverse variables. But the assignment variables have not been well investigated until recently.

The writing assignment usually gives the topic and describes the writing task. It may give the audience which the writer addresses and the format in which the student is to construct the response. The writing assignment variables may have a significant influence on writing performance. Unless these variables are controlled carefully, it can severely interfere with the obtained information about the student's writing performance. For example, different discourse modes with a diverse purpose and audience may require different organizational strategies and produce different levels of writing performance. Different topics may affect the quality of the writing sample because the topic-specific knowledge may vary from a student (or group of students) to another. Thus, one of the first and most important steps for writing assessment is to determine the kind of writing tasks or assignments that are critical for the purpose of the assessment.

Problem

In fact, research during the past decade has provided important evidence concerning the effect of several variables on the quality of writing. However, relatively little research has been done to explain the effect of assignment variables on the quality of writing compared to scoring criteria variables or rater consistency variables. As Breland (1983) reported in his review article on the direct assessment of writing skill, the history of direct writing skill assessment has been dominated by the issue of reliability. Much research

has been conducted about the rating inconsistency in the evaluation of writing samples. However, there has been little examination of the limitations in the degree of content sampling. Many writing researchers have assumed that writing performance has varied significantly according to the topic assigned and the student's knowledge and interest about the topic. But writing researchers have paid little attention to doing the empirical research necessary to lay the ground work for their assumption. Thus, the need for solid research about the relationship between topic variables and the quality of writing has recently been emphasized (Brossell, 1983; Greenberg, 1986; Hoetker, 1982a; Meridith and Williams, 1984; Purves, Soter, Takala and Vahapassi, 1984).

Currently there are two major suggestions concerned with the relationship between the structure of the writing assignment and the performance.

The first suggestion is that the purpose of discourse in which writers are asked to write produces different levels of writing performance. This suggestion is based on the belief that different purposes for discourse entails different thinking processes and results in distinctive stylistic features and organizational patterns; therefore, the writing skill required in accomplishing one rhetorical purpose does not necessarily mean that writing skill will be equally effective for a different rhetorical purpose. This means that different people may perform better with some rhetorical purposes than with others. In addition, a writing assessment which does not account for the purpose of the discourse variable is probably limited in validity (Odell, Cooper, and Courts, 1978).

The second suggestion is that the amount and specificity of information about the rhetorical context produces different qualities of writing. This suggestion is generally based on the idea that such information and specification allows writers a better chance to do their best work under test conditions, and it also enables writers to understand more fully the demands of the writing task. Consequently, writers produce a more coherent composition under those conditions.

Recently several studies have tried to determine the effect on writing of different discourse purposes, such as reference discourse, persuasive discourse, expressive discourse, and literary discourse. They have also been interested in the effect of the discourse mode, such as narration, description, exposition and argumentation. But this line of research has several inherent problems. One of the problems is that the characteristics of discourse purpose or discourse mode are uncertain. There is no firm agreement among discourse theories about the way in which discourse mode or discourse purpose interacts with the actual writing process.

Another problem is that current research does not consider the type or degree of content information given in the assignment. The type or degree of content information may lead to different levels of task complexity. Complexity levels of a writing task may not always reside in the general and broad distinction of discourse mode or discourse purpose because the same discourse mode or discourse purpose may have different levels of task complexity. For example, in the case of the expository discourse mode, the writer may be required to explain a simple fact or concept, using such techniques as comparison, classification, and cause-effect. At the opposite extreme, a writer may be required to write about a highly complex and abstract idea or a concept using the same expository discourse mode based on the type of content information given in the assignment.

The other problem is that most research uses general impression (holistic) scoring to investigate the effect of discourse purpose or mode on writing performance, but the general impression score provides no information about which element of the writing skill is affected by the given assignment variable. In his review article, Hoetker (1982a) pointed out the research problems of writing assignment variables:

We know that "mode of discourse" is somehow related to observable differences among student essays, but we don't know how mode may interact with variables involving students or subject matters.... We know that supplying a full rhetorical context should assist student writers, but we are not sure how to supply it so that it does indeed assist. (p. 389)

Purpose

Based on these problems, this study attempts to investigate an important characteristic of assignment variables: type of content information provided in the assignment (general vs. specific). The interest in the type of information given in the assignment is derived from the fact that the assignment variables have multidimensional characteristics and that current empirical research does not pay attention to the type of content information or type of question in the writing assignment, although a few researchers have conducted research on the type or degree of rhetorical information provided in the assignment.

Several researchers (Bridgeman and Carlson, 1984; Purves et al., 1984; Vahapassi, 1982) have clarified the multidimensional characteristics of the assignment variables such as cognitive processing demand level, task complexity level, and personal involvement level. In addition, the result of the research on conceptual elaboration of language processing suggests that the organization of prose is linked to its subsequent representation in memory; a writing task that requires a writer to draw more extensively upon previous knowledge produces more extensive conceptual elaboration for the task.

Thus, the purpose of this study is to determine whether two different assignments regarding the type of information (general vs. specific) will elicit a different quality of writing from a sample of college-level students based on the three evaluation measures: syntactic characteristic measures, analytic scores, and holistic scores.

More specifically, this study attempts to directly investigate

(a) whether the different assignments used by college-level students affect performance in the writing assessment,

(b) whether the assignment variables interact with different writer groups such as a student's major field and native language on the performance of writing,

(c) whether the different scoring scales leads to different interpretations of the effect of the assignment variables on the writing performance.

Limitations

For the purpose of this study, the limitation of the three main variables (topic, writer group, and performance) is necessary because an unlimited number of variations can occur in the actual writing assessment. Thus in this study, two types of topics (general content information provided in the assignment vs. narrowly defined specific content information provided in the assignment), three kinds of scoring scales (analytic scale, holistic scale, and syntactic characteristic measures), and two distinctive groups of writers (two different native language groups and two different major groups) will be used to investigate these relationships. A more detailed scheme of these criteria is described in the later section.

This study may have several limitations. It investigates the confounding effects of the task complexity and the type of information provided in the assignment. Therefore, the results of this study cannot discriminate between the task complexity effect and the information type effect. In addition, this study does not take into account the topic interest variables and topic specific knowledge variables. It is assumed that, with consideration of the Educational Testing Service (ETS) guideline for designing a writing assignment (Breland and Jones, 1982), the two variables are well controlled in the two writing assignments for the following reasons: First, each writing assignment used in this study presents a topic of the widest possible general interest to a diverse population. Second, the two assignments are brief and clear. Third, the vocabulary and the concepts presented in the assignment are not too difficult for ordinary college students to understand immediately. The audience variable is also not taken into account in this study because the audience is not specified in the two writing assignments. Thus, the result of this study may provide only limited information for the control of diverse variability in the writing assignment.

Despite these limitation, the results of this study may provide useful information for the development of a writing assignments, an important requirement for the valid assessment of writing performance. It may also provide useful information in answering the following controversial questions: Is success with one sort of writing task significantly different from success with another kind? Should narrowly defined specific information be provided for students so that they will achieve their best performance in the writing assignment? Should more than one writing sample or one scoring method be used as a valid and reliable test of student performance?

Hypotheses

In order to investigate all of the above relationships, the following null hypotheses are proposed:

(a) Based on the scores of analytic, general impression, and syntactic maturity scales, there will be no significant difference between the assignment with general information and the assignment with specific information on the performance of writing.

(b) There will be no significant interaction effects between the topic variable and the writer groups on the performance of writing. For example, both the nonnative speakers of English and the native speakers of English may be equally affected by the two different types of assignments. And both the hard-science major group the social-science major group may be equally affected by the two different types of assignments.

(c) There will be no significant correlations among the holistic scores, the analytic scores, and the syntactic characteristic measures.

CHAPTER II

REVIEW OF RELATED RESEARCH

In conjunction with the purpose of this study, this chapter deals mainly with studies on writing assignment problems in the assessment of writing performance and then briefly reviews major studies on the scoring methods in the writing assessment. The studies of the writing assignment problems are reviewed in regard to the theoretical perspectives and empirical studies. The review of the studies on the scoring methods focuses on the analytic and holistic scoring with consideration of the reliability problems.

Theoretical Perspectives on the Writing Assignment

The Role of the Writing Assignment

Current models of writing emphasize a role for the writing assignment in the writing process and product (Beaugrande, 1984; Bereiter, 1980; Bruce, Collins, Rubin, & Gentner 1983; Flower and Hayes, 1977, 1982, 1984; Gould 1978, 1980; Hayes and Flower, 1980, 1983; Martlew, 1983; M osenthal, 1983 etc.). For example, according to Hays and Flower (1980), the act of writing is divided into three major parts: the task environment, the writer's long term memory, and the actual writing process. According to this model, the major element of the task environment is the writing assignment, which includes a description of topic and intended audience. The information given in the writing assignment, with the writer's long-term memory, contributes simultaneously and interactively to the several subprocesses of writing, such as planning, translating, and reviewing. The process model of writing indicates that the writing assignment plays

an important role in the writer's ability to establish both global goals and local goals or to create the ideas and structure in the process of actual writing. According to Martlew's (1983) model, the writer's goal and purpose, which includes topic, reader (audience), discourse mode, and style, interact with the other writing processes such as memory, cognitive awareness, cognitive procedures, plans, linguistic expression, and the production of text. The interaction among the several parts in the model affects the resulting text in relation to achieving the writer's purpose and goal in respect to audience, topic, and mode of discourse.

This interactive view of the information processing approach was more fully investigated in the area of reading comprehension research (Just and Capenter, 1980; Kintch and van Dijk, 1978; Mitchell, 1982; Ruddell and Speaker, 1986; Rumelhart, 1977; Taylor and Taylor, 1983). Although the reading process is not exactly the inverse of the writing process, both reading and writing have essentially similar processes of meaning construction (Tierney and Pearson, 1983). The theory of interactive information processing in reading holds that the processing at different levels occurs simultaneously with influence passing in both bottom-up and top-down direction.

The theory conceptualizes reading not as a linear progression from sensory impression to memory nor from the reader's prior knowledge to understanding of text but as an interaction of both operating in parallel line. For example, according to Ruddell and Speaker (1986), the reading process incorporates four interactive components: reader environment, knowledge utilization and control, declarative and procedural knowledge, and reader product. Among the four components, the reader environment is closely related to the task environment of the writing process model. The reader environment components includes textual features (linguistic structure of the text), conversational features (message form, content, addresser, audience, tone, and manner), and instructional features (the structures of

learning objectives). The elements of the reader environment component interact with knowledge utilization and fmally interact with the reader's construction or representation of the text's meaning.

The interactive reading process model indicates that the writing assignment could play an important role in the production of the written text similar to the way the reading task environment plays a significant role in reading comprehension.

Elements of the Writing Assignment

The elements of writing assignments have been analyzed by several researchers in the context of the assessment of student performance on writing. Braddock, Lloyd-Jones, and Schoer (1963) identified the four elements of the writing assignment variables: topic, mode of discourse, time afforded for writing, and the examination situation. They suggested that variation in each of the four assignment elements would have significant effects on the quality of writing. Emig (1971) pointed out several internal aspects of the writing assignment that might influence the student's writing process and product: registers (the field of discourse, the written mode, and the tenor), the linguistic formulation of the assignment, the length, the purpose, the audience, the deadline, the amenities, and the treatment of written outcome. According to Emig, the general nature of the task (particularly the registers specified), the linguistic formulation of the assignment, the student's comprehension of the task, the student's ability to enact the task, and the student's motivation can affect the process and production of the writing.

More recently, Odell (1981) suggested that the purpose of discourse, audience, and discourse form were important elements of the writing assignment. According to his re-

search, the act of writing includes different kinds of modes or forms (dialogue, letter, memo, report, essay, and short story, etc.), diverse purposes (expressive, informative, and persuasive, etc.), and different audiences (self; well known other, and distant other, etc.). These different modes, purposes, and audiences require that students use different types of language, thought patterns, organizational strategies, and types or amount of information. He emphasized that the writing assignment should be explicit about the rhetorical purpose, the characteristics of the audience to be addressed, and the form. In his review article on the direct assessment of writing skill, Breland (1983) pointed out several elements of writing assignment variables: types of stimuli (written, pictorial, and aural, etc.), modes of discourse (narrative, expressive, and argumentative, etc.), context for writing (time allowed and reference material, etc.), the purpose of writing and the audience. He suggested that a combination of these assignment variables can produce extensive domains of writing samples, and, therefore, the use of only a limited sample from the domain may result in other errors in the writing assessment, in addition to errors resulting from rating inconsistency.

Based on the result of their work with the international study of written composition, Purves et al. (1984) developed a domain-reference system of writing assignments for the purpose of describing and classifying current writing assignments and also for creating writing assignments. The system provided a detailed description of the writing assignment, which included fifteen elements: instruction, stimulus, cognitive demand, purpose, role, audience, content, rhetorical specification, length, format, time, draft, and criteria. This system covered nearly all the elements of the writing assignments, which were suggested by previous studies.

The combination of these elements may affect students' writing performance. However, it may be impossible to investigate the effects of these combinations on writing performance. For this reason, several researchers have raised major questions about reliable and valid assessments of students' writing ability. For example, Odell, Cooper, and Courts (1978) raised the following questions: Might one's skill with one sort of discourse be significantly different from one's skill with other types of discourse? Should researchers provide a full rhetorical context (information about subject, audience, and purpose) in the writing assignment? How should researchers frame a writing task to obtain the best possible work from students? Odell (1979) raised other questions: Do different types of writing tasks elicit different kinds of writing performance from students? Does one writing task elicit a greater number of words or longer T-units and free 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?

In the context of the content and structure of the writing assignment, Hoetker (1982b) raised the following questions: To what extent should the topic supply context for the essay? What sort of discourse should the topic elicit? To what extent should the elements of the rhetorical situation be specified? What relationships are there between variations in the content and structure of the topic and the quality of the essay? Research on the above questions may provide new understanding of written production or the composing process and also may contribute toward the construction of valid and reliable assessment procedures. However, few empirical studies exist to provide answers for the above questions.

With consideration of the assignment elements identified, and the questions raised by several researchers, the elements of assignment variables can be classified into two major categories: (a) the type or extent of content information provided in the assign-

ment, (b) the type or extent of rhetorical information provided in the assignment. The first category may include those elements such as instruction, stimulus, content, topic familiarity, and information load. The second category may include those elements such as discourse purpose, audience, discourse mode, and discourse form.

Major Dimensions of the Writing Assignment

To synthesize and simplify those complex elements of the writing assignment, several researchers have provided simple models from two different perspectives. One approach is based on discourse theory or rhetoric, and another approach is based on empirical examination of the assignment variables in the writing assessment situation.

Discourse theorists suggest that multi dimensional aspects of the discourse model could provide important information for designing writing assignments or instructional programs in writing. Traditional discourse theory usually classifies the discourse types into a one dimensional category with the name of discourse mode (Bain, 1890; Brooks and Warren, 1972). The discourse mode usually includes four types of discourse: description, narration, exposition, and argument. However, Moffett (1968) distinguished two dimensions: the audience dimension (I-YOU relation) and the discourse mode dimension (I-IT relation). In the first dimension, Moffett displayed the kinds of discourse in a continuum of time abstraction: for example, the drama records what is happening, the narration reports what happened, and exposition generalizes about what may happen. On the second dimension, Moffett displayed discourse in a continuum of distance between author and audience, such as interior dialogue, conversation, public narration, and public generalization/inference. A similar line of classification was provided by Britton, Burges, Martin, Mcleod, and Rosen (1975). They developed a two-dimensional

model of school writing: sense of audience dimension and dominant function of discourse dimension. They divided the audience dimension into self, teacher (examiner), wider audience (known), unknown audience, and additional categories. The function dimension was classified into three main categories: transactional, expressive, and poetic. The transactional writing was subdivided into informative (e.g., record, report, generalized narrative, analogic, and tautologic) and connative (e.g., regulative and persuasive).

While the above two classifications emphasized the discourse purpose dimension and the audience dimension, Kinneavy (1971, 1980) classified discourse of writing into the discourse aim dimension and discourse mode dimension based on his communication triangle (encoder, decoder, and reality). The discourse mode dimension included narration, description, evaluation, and classification. The discourse aim dimension included expressive, referential, literary, and persuasive. As with the classification of Kinneavy, D'Angelo (1976, 1980) suggested two dimensions of discourse: purpose/aim as one dimension, and mode as the other dimension. He classified the purpose/aim dimension into informative, persuasive, literary, and expressive discourse. The mode dimension, which referred both to a way of getting ideas to write about and to the manner of developing those ideas, was categorized into the static mode (topic) and progressive mode (topic). The static mode included identification, analysis, description, classification, exemplication, definition, and comparison/contrast; and the progressive mode included narration, process, and cause/effect. According to D'Angelo, all these categories can be used to direct the search for ideas or the arrangement of those ideas into some orderly pattern.

From a different point of view, Brewer (1981) classified discourse into two dimensions: discourse type and discourse force. He suggested three basic types of discourse:

description, narration, and exposition. According to Brewer, the description is based on the visual and spatial cognitive structure, the narration is based on the mental representation of a series of temporary occurring events, and the exposition is based on the mental process of such things as induction, classification, and comparison. The discourse force dimension was classified into purposes: to inform, to persuade, to entertain, and the literary-aesthetic. In each of the four discourse forces, a writer has a different aim or intention: for example, the informative discourse focuses on giving information about something, but the persuasive discourse focuses on convincing or persuading the reader to take a particular course of action or to adopt a particular set of ideas.

While the above researchers' classification scheme is mainly focused on the functions/aims of discourse, types/kinds/modes of discourse, and the writer-reader relationships (audience), Vahapassi (1982), Purves et al. (1984) and Bridgeman and Carlson (1984) suggest major dimensions of assignment variables in regard to cognitive processing demand levels, task complexity levels, or information load levels. Vahapassi (1982) and Purves et al. (1984) developed two major dimensions of writing assignment: level of cognitive processing demand as one dimension, and dominant intention/purpose as the other dimension. The model was based on an examination of the assignments used for an international study of written composition and on a review of the various theoretical perspectives on composition. The cognitive processing dimension includes three levels of cognitive process: reproduce, organize or reorganize, and invent or generate; and each processing level is matched with a primary mode of discourse and content. The dominant intention or purpose dimension includes five cases of discourse aims: to learn (metalingual), to convey emotion or feeling (emotive), to inform (referential), to convince or persuade (connative), to entertain (poetic). Each discourse aim was then matched to a different audience.

Based on a survey of academic writing tasks, Bridgeman and Carlson (1984) reported a two dimensional scheme of the writing assignment variables. Their multidimensional scaling of the topic types shows a two dimensional space defined by a complexity dimension and a personal involvement dimension. In the case of the complexity dimension, a certain type of assignment may require more direct and fewer analytic thinking skills; whereas, another type of assignment may require more complex and diverse thinking skills. In the case of the personal involvement dimension, a certain type of assignment may provide some information that the writer uses in completing the task requirement, while another type of assignment does not provide enough information and may demand that the writer bring a high degree of personal knowledge and experience to the writing task.

Summary of the Theoretical Research

Theoretical research on the writing assignment was reviewed in regard to the role, the elements, and the major dimensions. According to the current interactive models of writing, the writing assignment (or writing task) has been regarded as an important part of the model because the writing task or the writing assignment can play an important role for the writer to establish both global goals and local goals in the process of writing. According to the suggestions made by several writing researchers, the elements of the writing assignment can be classified in two major categories: (a) the type or extent of rhetorical information provided in the assignment and (b) the type or extent of content information provided in the assignment. The rhetorical information includes those elements such as discourse mode, discourse purpose, and audience. The content information includes those elements such as instruction, stimulus, content, topic familiarity, and information load. The multidimensional characteristics of the writing assignment were

suggested by two different perspectives: the view based on discourse theory or rhetoric and the view based on empirical examination or survey in the context of writing assessment. In the view of the discourse theory or rhetoric, the two major dimensions are discourse purpose and audience or discourse purpose and discourse mode. In the view of the empirical examination or survey, the two major dimensions are discourse purpose and cognitive processing demand or task complexity and personal involvement. Taken as whole, the above review clearly shows that the control of the writing assignment variables is not simple and straightforward in the writing assessment situation.

Empirical Research

Research on writing assignments has been approached from two major viewpoints. One approach is to focus on the single element of rhetorical information (such as discourse mode, discourse purpose, and audience) in which students are required to write and then to explain what discourse mode, purpose, or audience results in higher performance. The second approach is to focus on the degrees or characteristics of information provided in the assignment and to try to explain whether certain degree or characteristics of the information provided in the assignment help or hinder students' writing performance. The second approach has looked at the following three aspects of the writing assignment: topic familiarity, degree of rhetorical information, degree of content information.

Research on Discourse Mode, Discourse Purpose, and Audience

Discourse mode. Since Braddock et al. (1963) called attention to the effect of the mode of discourse variable on student writing performance, a number of studies have

been conducted. Several other researchers found different effects of discourse modes on writing performance.

Veal and Tillman (1971) examined the mode of discourse variation in the assessment of children's writing. The tests, consisting of one topic in each mode of discourse (argument, description, explanation, and narration), were administered to two classes of second, fourth, and sixth grade children. The compositions were rated with a 7-point rating scale for each of the four topics. The results show that the largest overall differences were obtained between second and sixth grade papers for the expository mode, and these differences were followed by narrative, descriptive, and argumentative modes. In the second grade, the differences between modes were not significant; but, in the fourth grade, the scores of descriptive and expository modes were significantly higher than the scores of argumentative modes. There was a pattern of increasing complexity in the relationship between modes: little increasing quality in the argumentative mode relative to the greater increasing quality in the narrative and descriptive modes.

Quellmalz and Capell (1979) studied the relationship of writing scores of high school students when students were tested on writing tasks by either writing the same discourse mode or two different discourse modes. On all five subscales and on total essay scores, the narrative ratings were lower than the expository ratings, and the relationship between a student's two essay scores on general impression, organization, or total score was higher when they wrote in the same discourse mode than when they wrote in different discourse modes. The study suggests that the result may be due to three causes: (a) the different curricular emphasis given to narrative and expository writing in the schools, (b) the student's lack of knowledge at the personal-experience level to deal with

the narrative topics, or (c) the raters' tendency to score expository essays more highly than narrative essays.

Crowhurst (1980) studied the relationship between syntactic complexity and the quality ratings of narrative and argumentative writing of students in grades 6, 10, and 12. The results showed that the writings in the narration mode received higher scores at three grade levels than did the writings in the argument mode. In addition, it was found that in the narrative mode, a high degree of syntactic complexity was not associated with higher quality scores; whereas in the argument mode, a high degree of syntactic complexity was associated with higher quality ratings.

Reed, Burton, and Kelly (1985) investigated the effects of writing ability and mode of discourse on cognitive capacity engagement across the three stages of the composing process. They found that writing ability affected cognitive capacity engagement across discourse modes. High level writers were least engaged when writing descriptive essays but were most engaged when writing argumentative essays. Average writers, on the other hand, were most engaged when writing descriptive essays but were least engaged when writing narrative essays. This study suggests that the level of development of writing to the related schemata is indicated by the engagement required and the quality of writing produced.

In summary, the above research on discourse modes shows the following three major points. First, student writing performance was significantly affected by different discourse modes. Second, the performance in the argumentative discourse mode are lowest among the different discourse modes. Third, the discourse mode effects on student writing performances are stronger for the elementary school students than for the high

school students. The second and the third point may be related to the developmental trend of writing skill because the argumentative mode requires students to use a higher level of logical reasoning skill, and the skill may not be developed in elementary-school level students. One of the problems with the research on discourse mode effect is derived from the fact that the different discourse modes can easily be blended in a piece of writing because the single essay is usually composed of smaller units of different discourse modes. Therefore, the results of studies in discourse-mode effect on writing performance should be interpreted with caution.

Discourse purpose. Recent discourse theory gives a great deal of attention to discourse purpose (Britton et al., 1975; D'Angelo, 1976; Kinneavy, 1971; Moffett, 1968). According to Odell et al. (1978), discourse modes are important only as the means by which a writer accomplishes a given discourse purpose, and skill in different discourse modes is of little use unless that skill serves some larger discourse purpose. However, a few studies have been conducted for the effect of discourse purpose on student writing performance.

Prater (1982) conducted a study to compare the writing performance of tenth grade students across different discourse purposes. The subjects were eighty students whose writing skills had been classified as either master or nonmaster, based on their scores on the writing section of the Texas Assessment of Basic Skills. These students were assigned three writing tasks, each of which elicited a different discourse purpose: expressive, explanatory, and persuasive. Prater found that the students' writing performance varied across discourse purposes. The expressive writing samples received the highest scores, followed by persuasive writing samples, but no difference was found between the per-

formance of master and nonmaster students. He suggested that this result points out the inadequacy of using a single writing sample to judge writing competency.

Prater and Padia (1983) investigated the effects of discourse purpose on writing performance in grades four and six. The subjects of their study were 70 fourth grade and 70 sixth grade students. All students wrote on each of three writing tasks (expressive, explanatory, and persuasive), and each essay was scored using a four point holistic scale. According to the results, the students' expressive writing received the highest ratings, followed by the explanatory, and then the persuasive writing. The results indicated that the expressive writing was particularly well suited to the elementary age writer because of the ego-centric nature of the writing task; whereas the persuasive writing task was difficult for this age group because the task required students to address an unseen and rhetorically distant audience. They suggested that a sample of student writing focussing on one discourse purpose could not provide an accurate measure of overall writing proficiency.

Cadman (1985) investigated the effects of scoring methods, topics, and purpose on grade 12 students' writing scores. Each student wrote on two transactional topics and two expressive topics in counter-balanced order on four different occasions in the English classroom. The result showed that writing for different purposes and on different topics produced different scores for individuals scored by primary trait scoring. The study suggests that when assessing writing for group purposes, such as large scale assessment or research, the topics should elicit the same discourse purpose and should be equated in the domain of content and distance from the writer; whereas a writing assessment for the purpose of achievement, admission, placement, or competency should evaluate writing based on more than one discourse purpose.

In summary, the above empirical research on discourse purposes show that elementary and high school student performance on writing is significantly affected by different discourse purposes, and that the student performances are lowest in the persuasive discourse and highest in the expressive discourse. In addition, the results of the above research indicate that expressive discourse is well suited for the elementary school students. One of the problems in the above research on discourse purpose effect is derived from the fact that the discourse purpose called for by the writing assignment does not control uniformly the discourse of student writing. As Pianko (1979) reported, students usually produce writing in the discourse purpose that they believe it calls for, instead of the discourse purpose that the assignment calls for.

Audience. Most empirical research on audience adaptation has focused on the writer's ability to differentiate audience by asking students to write the same message for audiences with different characteristics, such as differences in age, intimacy, or power.

Crowhurst and Piche (1979) studied the effects of audience on syntactic complexity. They asked sixth and tenth grade students to write three different modes of writing (narration, description, and argument) to each of two audiences: a best friend and a teacher. They found that, at sixth grade, the syntactic complexity of the two papers written for the two audiences was not significantly different; however, at tenth grade, the papers written for the teacher audience were syntactically more complex than for the best friend audience. In addition, the audience differences were most clearly evident in the mode of argument because the argument contributed to a greater difference in syntactic complexity for the different audiences than the description or narration mode.

Rubin and Piche (1979) investigated the effects of the audience adaptation ability on persuasive writing written by students in grades four, eight, twelve, and adults. The audience variables were someone the writer knows very well, someone the writer does not know very well, and generalized other. According to the results of their content analysis, only the expert adult writers showed substantial differences in the strategies they used for different target audiences. The expert adults used a large number of appeals, a large variety of appeal types, and more advanced categories of appeals than the grade school students.

While these researchers investigated the audience effect based on the syntactic complexity or the content analysis of the essay, several researchers have focused on the relationship between audience adaptation ability (or social cognitive ability) and the quality of the writing. For example, Rubin, Piche, Michlin, and Johnson (1984) examined the contribution of measured audience adaptation ability to the overall quality of writing in fourth grade students' narratives. The results showed that the audience adaptation ability accounted for 36.3 % of the variance in quality rating, indicating that the social cognitive ability contributed substantially to the overall quality of the written compositions. Burleson and Rowan (1985) examined the relationship between social cognitive ability and narrative writing skill. The results showed that there was no re- lationship between social cognitive ability and narrative writing.

In addition, Rubin and Rafoth (1986) examined the relationship between audience adaptation ability (social cognition) and writing quality. The result indicated that the audience adaptation ability played a more significant role in writing a persuasive message to a relatively determinate audience. According to their canonical correlation analysis, the audience adaptation ability alone predicted 26 % of the variance in the

quality rating of the persuasive writing. Kroll (1985) examined the relationship between social cognitive ability and writing performance in 49 nine-year-old children's composition written for four different discourse purposes (narrative, expressive, referential, and persuasive). In the study, the children produced an oral task and completed a social cognition task. The results showed that social cognitive ability was most strongly related to the oral task and weakly related to the performance on the four writing tasks.

Rafoth (1985) examined the difference between the essays of proficient and nonproficient college freshman writers in the aspect of the audience adaptation. In the study, both proficient and nonproficient writers wrote a persuasive essay in two types of audience presentation conditions: content rich and content poor. According to the results, proficient writers were better able to make use of additional audience information than nonproficient writers, and in the absence of additional audience information, proficient and nonproficient writers were similar in audience adaptation. Along with other researchers, such as Flower and Hayes (1982), Sommers (1980), and Rubin et al. (1984), Rafoth's study confirmed the position that proficient writers use more adaptation strategies and made more specific links between their essays and available information about the audience than nonproficient writers.

Taken as a whole, the above studies on audience adaptation ability generally suggest the following three major points: (a) audience adaptation ability is most highly correlated with the quality of the persuasive writing, (b) in the case of the other writing tasks (such as descriptive, narrative, and expository), the results of research on the relationship are mixed, and (c) the audience adaptation ability is not developed in elementary level or high school level students.

Research on Degree or Type of Information

While most empirical research on the writing assignment has mainly focussed on a single element of the rhetorical information (such as discourse mode, purpose, and audience), more recently, some researchers have focused their attention on the degree of or characteristics of information provided in the assignment in regard to the information load levels and the task complexity levels.

Degree of rhetorical information. Brossell (1983) investigated the effects of rhetorical specification in essay examination topics. The study used three different levels of essay topics: level 1--low information load; level 2--moderate information load; and level 3--high information load. The level 1 topic presented a briefly stated subject, leaving the writer free to make decisions about audience, purpose, and speaker without guidance of any kind. The level 2 topic presented an introductory statement about a subject and the task but omitting information about purpose, audience, and speaker. The level 3 topic presented a hypothetical situation, or scenario, requiring the writer to state personal , views on a topic in a full rhetorical context. The results showed that different levels of essay topics did not affect the quality of student writing, and essays written at full rhetorical specification had a lower mean score and a greater mean length than essays written at low or moderate rhetorical specification.

Bates (1985) examined the writing performance of college freshmen and its relationship to writing attitude, topic knowledge, and specified writing goal. In the study, all subjects received a similar writing task, but experimental conditions varied according to how fully the writing goals were specified and whether or not additional topic-related knowledge was available in the form of extra reading material. The results showed that

the specified writing goal had a positive effect on writing quality rated by the general- impression scoring method, but that different topic knowledge did not have any significant effect on writing quality.

Taken as a whole, the above research reveals two major conclusions. First, the information load level of the writing assignment does not have a significant effect on student writing performance. Second, a writing assignment with the highest information load tends to produce loosely organized and low quality essays, although the difference in quality is not statistically significant from the assignments with low level or medium level information load. The results of the above two studies support Hoetker's (1982b) suggestion: "The more information students are given--the more language they have to process--the greater seems to be the opportunities for creative misreading and simple confusion." (p. 11)

Degree of task complexity. Greenberg (1981) studied the effect of variation in essay questions on the writing performance She attempted to determine whether variation in essay questions would increase the writer's performance. The variation of essay questions was manipulated into cognitive-demand levels and experiencial-demand levels. Cognitive-demand levels refer to the degree of information and the number of strategies which an essay question provides, and experiencial-demand levels refer to the degree of personal experience that the essay question demands. The categories of writing performance were (a) overall writing quality, (b) syntactic complexity, (c) sentence control errors, and (d) essay length. Although Greenberg hypothesized that there would be a significant interaction between high cognitive and low experiencial-demand levels of the essay questions with low cognitive and high experiencial-demand levels of questions in the quality of writing, the result showed that changes in the essay question did not affect

students' writing performance. One major problem of the Greenberg's study is that the four writing tasks have the same degree of cognitive-demand level because the small changes in the wording of the writing assignments (otherwise similar in regard to information load and rhetorical information) may not lead to different degrees of cognitive- demand levels or personal-involvement levels.

Nold and Freedman (1977) attempted to determine what elements influences a teacher's rating of student writing. For the study, 22 college freshmen wrote four in-class essays on four different topics. The topics were all in the argumentative mode of discourse because variation in mode of discourse may have more effects than variation in topic on the quality of writing. Two of the four topics were "matched personal opinion" topics, and the other two topics were "compare quotation" topics. The results showed that the personal opinion argument topics yielded more writing, a greater variation in amount of response. In the compare argument topics, there was less writing, and less variation in amount of response.

Taken as a whole, the above two studies show that the task complexity effect is not significant when the complexity levels of the writing tasks being perceived by students as very similar tasks. Inaddition, the studies show that the different task-complexity levels affect the syntactic characteristics of the student essays.

Summary of the Empirical Research

The above research shows that there are significant discourse mode effects, discourse purpose effect, and audience effect on the quality of student writing. At both the elementary-school level and high-school level, the scores of papers written in the

argumentative discourse mode are significantly lower than the scores of those papers written in the descriptive, expository, and narrative discourse modes. In the case of discourse purpose, the scores of persuasive discourse papers are significantly lower than the scores of expressive and explanatory discourse papers. In addition, the audience effect on the student writing performance is most significant in the persuasive discourse papers. These results indicate that the discourse mode, purpose, and audience effects are strongly related to the developmental trend of writing skill because the argumentative discourse mode and the persuasive discourse purpose require students to use high levels of logical reasoning skills or social cognitive abilities. These skills may not be fully developed at the elementary-school level or the high-school level.

The above studies on discourse mode and purpose are based on samples of elementary and high school student writing. Therefore, further research is needed to investigate the effects of discourse modes and purpose on samples of college-level student writing. Furthermore, current research on discourse mode and purpose has several problems. One is that the researchers have not considered the degree or characteristics of the information provided in the writing assignment. The other problem is that the discourse mode and purpose have overlapping traits in the overall text of a single essay.

Research on the degree of information provided in the writing assignment shows two major results. First, the information-load level of the assignment does not have a significant effect on the student writing performance, although the assignment with the most fully specified information load tends to yield loosely organized and low quality essays. Second, the task-complexity level of the writing assignment does not have a significant effect when the student perceives the writing tasks as not requiring distinctively

different levels of task complexity in spite of the different wordings used in the writing assignments.

Research on Scoring Method and Reliability

As Hirsch (1977) suggested, finding a reliable and valid assessment is the "single most important snag to practical progress in composition teaching and research" (p. 176) because the assessment of student writing performance can provide important information for carrying out instructional or administrative functions and for carrying out research in composition. Many researchers have devoted their efforts to ford a scoring method that is both reliable and valid. In addition, the scoring method might play an important role in investigating writing assignment effects on writing performance because writing assignment effects may be interpreted differently, according to different scoring methods. Therefore, in the investigation of the writing assignment effects, one should consider the scoring method effects in conjunction with the writing assignment effects. This section briefly reviews major research on scoring methods with consideration of reliability issues in writing assessment.

At the present time, there exists a number of different scoring methods for assessing student writing performance. Breland (1983) classified the scoring methods into eight categories: holistic, focused holistic, analytic, atomistic, primary trait, syntactic, communicative effectiveness, and automated scoring. Among these different scoring methods, the holistic scoring method and the analytic scoring method are more widely used than the other methods. The syntactic characteristic measure is often used for special purposes such as deciding the syntactic complexity levels or syntactic maturity

levels of student essays (Myers, 1980; Meridith and Williams, 1984). Because of the purpose of this study, this review deals only with two scoring methods--analytic and holistic. Research on the syntactic characteristic measures is discussed in the next two chapters.

Analytic Scoring Method

The analytic scoring method focuses on several important elements or characteristics of writing skill (Cooper, 1977; Odell, 1981; Mullis, 1984). The most fundamental task in the analytic scoring method is to establish the criteria of important components or characteristics of writing skill. The scoring criteria must be consistent with judgments about good writing in the literate society at large (Hirsch, 1977).

The most well known analytic scoring criteria were established by Diederich, French, and Carlton (1961). The criteria were based on a factor analysis of writing samples scored by sixty expert readers representing a variety of academic fields. The factors derived were ideas, form, flavor, mechanics, and wording. In a later version (Diederich, 1966, 1974), the mechanic criterion was further divided into usage, punctuation, spelling, and handwriting. According to the scoring method, each factor is rated on a scale from 5 to 1, and the idea and form factors receive a double weighting. Diederich (1974) reported that when student essays were rated by the analytic scoring method on the basis of two reading per student, the resulting reading reliability was .80. Their factor analysis shows that the five factors represent five different school of thought in the rating of essays. One characteristic of the Diederich analytic scoring method is that the system is based on the judgment of different raters (representing several different academic disciplines) without considering different writing tasks (or assignment) variables.

Other analytic scoring criteria were identified by Remodino (1959). To clarify the distinguishable qualities of student essays, seventeen categories, derived from a survey given to twenty teachers of literary subjects, were assessed in 230 compositions. Remondino found that the seventeen elements of writing quality could be reduced to four factors: graphic representation, language usage, content and arrangement, and personal aspects of the content. Although he did not clearly establish the analytic scoring criteria, he pointed out that not all of the factors carried the same weight, from an educational point of view.

While the above two analytic scoring criteria were established on the basis of different raters' judgments about good writing within a single literate society, a cross-national study of student achievement in written composition undertaken by the IEA, established another analytic scoring criteria, using an international sample of raters (Purves, 1984). In the IEA study, 26 elements of writing quality were identified, based on the international raters' comments on representative compositions from each rater's country. The IEA study classified those 26 elements into general criteria. For this purpose, six raters rated the international students' compositions for five different topics on the basis of 26 elements of writing quality.

According to the results of a factor analysis, based on the five raters' scores, four factors were generated: organization or structure, style, quality and scope of content, and personal affect. From the factors generated, the IEA study developed an analytic scoring scheme that included four analytic criteria: (a) quality of scope and content, (b) organization and presentation of content, (c) style and appropriateness of language (including tone), and (d) personal response of the reader. These four criteria could be important for any writing task. To these four main criteria, the IEA study added another four criteria, usually emphasized in the educational context: (a) mastery of grammatical convention, (b) mastery of orthographic convention, (c)

spelling, and (d) handwriting and neatness. The raters from fifteen countries, who participated in the IEA study found these criteria acceptable. One important characteristic of the IEA scoring system is that it has flexibility for different kinds of writing tasks because it provides a relatively efficient way to apply different subcriteria to different writing tasks.

The analytic scoring method can provide diagnostic information about the strengths and weaknesses of each piece of student writing (Mullis, 1984). In addition, the analytic scoring method leads to higher rater reliability compared with the holistic scoring method (Bauer, 1982; Breland, 1983; Coffman, 1971a, 1971b). However, only few reports exist on the reading reliability (sometimes called rater reliability) and score reliability using the analytic scores. The reading reliability reflects error variance associated with reading inconsistencies among readers, and the score reliability reflects both the error variance attributable to the reading inconsistencies and the error variance associated with sampling (different writing assignments). According to Breland (1983), the reported reading and score reliability estimates of the analytic score are as the following:

1. reading reliability estimates:
   - Coward (1952); .70 in one reading for one task.
   - Diederich (1974); .82 in two readings for one task.
   - Bauer (1982); .95 in three readings for one task.
   - Breland (1983); .67 in one reading for one task, and .80 in two readings for one task.
2. score reliability estimates:
   - Steele (1979); .82 in two readings for two tasks.
   - Quellmalz, Capell and Chou (1982); .83 in two readings for two tasks.

Holistic Scoring Method

The holistic scoring method is based on the view that a piece of writing is greater than its parts (Cooper, 1977; Hirsh and Harrington, 1981; Mullis, 1984). The holistic score provides information about the range of overall writing quality. In using the holistic scoring method, the raters simply make a single global judgment about each paper by deciding where the paper fits within the range of papers produced for the writing assignment.

The holistic scoring method was originally developed by the Educational Testing Service (ETS), and the scoring method was firmly established by a large-scale investigation conducted by Godshalk, Swinford, and Coffman (1966). One of the purposes of the investigation was to validate the use of the holistic scoring method as a direct measurement of writing skill. In their experimental study, conducted in 1962, holistic scores were generated on 646 essays for five different topics. The essays, written by eleventh-grade and twelfth-grade students, were independently scored by five different readers on the basis of the rater's global or holistic judgment. In another field trial study, conducted in 1962, a group of 145 readers reread two of the topics in order to assess the effects of reading on a larger number of readers using the four-point scale.

The results of these two studies yielded the following (Godshalk et al., 1966): First, the reliability of essay score is primarily a function of the number of different essays and the number of different readings included. For example, for five topics read by five readers, the reading reliability of the total score approximates .92 and the score reliability approximates .84. In contrast, for one topic read by one reader, the corresponding fig-

ures are .40 and .25 respectively. In addition, the reading and score reliabilities, based on the four-point scale, are higher than the reliabilities based on the three-point scale. Second, when an objective test of writing skill is evaluated against a reliable criterion of writing skill, the test is highly valid. Third, the most efficient predictor of a reliable direct measure of writing ability is one which combines an essay test with an objective test.

More recently, the holistic scoring method has been revised by the ETS (Mullis, 1984). According to Mullis, in the revised holistic scoring method, "standard papers are selected to represent the various score points, and raters are carefully trained to become calibrated to reach consensus." (p. 17) In addition, raters are trained to use the full range of scores available to approximate a normal distribution. The standard papers are representative of all the papers at a given scoring level. Hence, the standard papers should be typical because the papers are used in rater training, and later used as models to assist raters during scoring (Spandal and Stiggins, 1981). With the standard papers, sometimes brief guidelines or features are provided the raters for the purpose of describing general attributes of papers in each quality level. The primary purpose of the holistic scoring is to rank students according to overall writing proficiency, so, the scoring is rapid and efficient. However, the holistic score does not provide diagnostic information about the proficiencies or deficiencies in student writing.

Several researchers reported the reading and score reliability of the holistic score (Breland, 1983). The reported reading reliability estimates of the holistic score for one task are as follows:

- Coffman (1966); .39 for one reading, .56 for two readings, and .65 three readings.
- Coffman (1971b); .70 for two readings.

o Breland (1983); .54 for one reading, and .70 for two readings.

According to Breland (1983), the overall median estimates of the reading reliability estimates of the holistic score for one task are .64 in one reading, .70 in two readings, and .78 in three readings.

The reported score reliability estimates of the holistic score for two tasks are as follows:

o Coffman (1966); .42 for one reading, .55 for two readings, and .66 for three readings.
o Clemson (1978); .55 for two readings.
o Steele (1979); .58 for two readings, and .62 for three readings.
o Breland and Gaynor (4979); .51 for two readings.

According to Breland (1983), the overall median estimates of the score reliabilities of the holistic score are .53 in two readings for one task and .69 in two readings for two tasks. Breland (1983) proposed new estimates of reading and score reliabilities for various combination of tasks and reading per task on the basis of empirical data. His new reliability estimates are higher than those of Coffman (1966): the reading reliability for two readings .70 for one task; .81 for two tasks; .85 for three tasks; .88 for four tasks; and .90 for five tasks. The score reliability for two readings is .53 for one task, .70 for two tasks, .76 for three tasks, .81 for four tasks, and .88 for five tasks.

CHAPTER III
RESEARCH METHOD AND DESIGN

Research Method

The Sample of This Study

The original writing sample in this study consist of 424 essay written by 212 upper- college level students. The sample includes 96 native speakers and 116 nonnative speakers of English. In addition to the writing sample, the following information was available: the student's native language, major field, GRE scores, and holistic (general impression) score for two different topics. (Use of the data was granted by the Educational Testing Service (ETS)).

From the original ETS data, ninety six student writing samples were selected for this study--twenty-four students, native speakers of English, with a major in a hard science; twenty-four students, native speakers of English, with a social science major; twenty-four students, native speakers of Chinese, with a major in a hard science; and twenty-four students, native speakers of Chinese, with a social science major. Since each student wrote two essays on two different topics, the total number of writing samples for this study is 192.

Quality of the Essays

Three different kinds of writing skill scoring methods were used: analytic score, holistic score, and syntactic characteristic. The analytic scores were generated by four

raters using Purves's analytic scoring scheme (1985). The analytic scoring scale includes three main criteria and thirteen subcriteria:

A. Content and Thinking main criterion:

1. richness of information, 2. adequacy of Information, 3. relationship,

4. inferences, 5. synthesis, 6. evaluation, 7. alternatives.

B. Organization main criterion:

8. Framing, 9. Grouping, 10. Unity.

C. Style and Tone main criterion:

11. Objectivity, 12. Tentativeness, 13. metalanguage.

Based on the analytic scoring scale, all essays were rated by four raters with a great deal of knowledge and experience in the rating of English compositions. The raters initially met in two sessions for discuss criteria and to rate ten writing samples. Raters were able to use the comprehensive criteria information provided in the form of chart (detailed rating procedure; Soter, 1985). For the "Space" topic, all four raters rated all essays, whereas for the "Farm" topic, two raters rated the odd numbered essays, and another pairs of raters rated the even numbered essays. Finally, the quality of the "Farm" topic essay was rated by two different rater groups.

Therefore, to minimize the rater effect on the comparison of the two topics, the average scores of the two rater groups' rating for both the "Space" topic and the "Farm" topic were used in the investigation of the topic effect on students writing performance. The inter-rater reliability for the original writing samples shows that the analytic ratings between four raters were consistent (inter-rater reliability coeffeicient alpha for "Space" topic = .90, and for "Farm" topic = .84). Consequently, it is assumed that the rater variable does not significantly affect the quality measure of these writing samples when the average score of the

two rater groups' ratings for each writing sample is used as the quality measure of the writing sample.

The holistic scores were generated on a six-point scale for the original writing samples. The inter-rater reliability for the original writing samples shows that the holistic ratings between two raters were consistent (inter-rater reliability coeffeicient alpha for "Space" topic= .81, and for "Farm" topic = .87). In the holistic scoring method, the raters independently rated students' essays by evaluating whether the paper fits within the range of essays written for the given assignment. The raters were trained to read the essay quickly and to score it as a whole without considering the several dimensions of writing skill (Breland and Griswold, 1981).

In addition to the analytic scores and general impression scores, two kinds of syntactic characteristic measures--length of elaboration measures and syntactic complexity measures--were used as dependent variables. For the length of elaboration measures, the total number of words, total number of T-units, total number of free modifiers (openers, interrupters, and closers) were counted, using several reference materials (Christensen, 1968; Christensen and Christensen, 1978; Dixon, 1970; Hunt, 1970, 1977; O'Hare, 1973; Tibbetts and Tibbetts, 1984; Wolk, 1970). For the syntactic complexity measures, the mean T-unit length and the ratio of free modifiers (ratio of openers, interrupters, and closers) were counted in the course of the data analysis.

The T-unit, (a group of words which constitute a main clause in addition to all subordinate clause attached to it) and the ratio of free modifiers were used for determining syntactic complexity or maturity because it is widely accepted that the mean T-unit length increases with maturity and a mature style has a relatively high frequency of free modifiers, especially in the final position (Christensen, 1968, 1978; Hunt 1977). According to Wolk (1970), the initial free

modifier (opener) includes all words, phrases, and clauses that precede the noun phrase which serve as the subject. The medial free modifier (interrupter) is set off by punctuation and should occur neither initially nor finally. The final free modifier (closer) is set off by punctuation, and appears after the last word of the bound predicate.

Characteristics of the Essay Topic

Each student wrote two essays on two different topics. For the "Space" topic, students were required to compare and contrast the advantages of space exploration and to take a position. In the"Farm"topic,students were required to interpret therelationships among three graphs showing the change in farming patterns over a period of forty years in United States. These two topics are distinctive in regard to the type of information provided in the assignment (See Table 1).

According to Bridgeman and Carlson's (1984) topic type analysis scheme, a topic type, like compare and contrast plus take a position, is located in the high complexity and high personal involvement dimensional space, and the other topic, interpret graph or chart, is located in the low complexity and low personal involvement dimensional space. Bridgeman and Carlson suggest that the compare and contrast topic requires the student to generate two parallel lines of reasoning and to relate them to one another, but the interpret graph or chart topic supplies the material that the writer needs in order to complete the task. However, for the "Farm" topic of this study, students were required not only to interpret a single graph, but also to interpret the relationships among the three graphs, based on comparison and causation of the changing patterns in farm size, farm population, and number of farms. Therefore, the two topics of this study have similar levels of task complexity.

Table 1

Topic Comparison in Relation to the Elements of Writing Assignment

	"Space" topic	"Farm" topic
Instruction type	topic and content	topic and content
Stimulus (information type)	one paragraph with general information	one paragraph with specific information using graphs
Cognitive demand	invent/generate/evaluate	invent/generate/evaluate
Purpose	to convince/inform	to inform/convince
Audience	unspecified (general)	unspecified(general)
Mode	argument/exposition	argument/exposition

On the basis of this interpretation and the suggestions of the preceding research (Purves et. al, 1984; Bridgeman and Carlson, 1984; Vahapassi, 1982; Brossel, 1983; Greenberg, 1981), the two essay topics can be distinctively located in the hypothetical multidimensional space as shown in Figure 1.

Research Design

The three purposes of this study were to investigate (a) the topic effect on the student writing performance, (b) the topic and the writer group interaction effect on the writing performance, and (c) the correlations among the different measures of writing skill. For the first and second purpose, the repeated measure of analysis of variance method, based on the split-plot factorial design is used in this study. For the second purpose, the correlational analysis method is used.

Cognitive Processing Demand Level

high	low		
A		general	Information Type
B		specific	

(topic A; compare/contrast and take position)
(topic B; interpret graph and report)

Figure 1. A hypothetical multidimensional space.

Analysis of Variance

According to Kirk (1982), the split-plot factorial (SPF) design is appropriate when repeated measures are used in the analysis, and the SPF p.q design (p before the dot = between subject treatment level, and q after the dot = within subject treatment level) provides the great power associated with test of treatment B (in this study, the topic) and AB (in this study, the interaction of the writer group and the topic). Because the main interest of this study is in the topic effect and the interaction effect between the topic and the writer group, the SPF design may be appropriate for this study. For the analysis of variance, the following dependent variables and independent variables are used in this study.

Dependent variables. Three kinds of elaboration length measures, and three kinds of syntactic complexity measures, and four kinds of analytic score, and one kind of holistic score are used as dependent variables.

A. Syntactic characteristic measures: length of elaboration
   1. total number of words
   2. total number of T-units
   3. total number of words in free modifiers
B. Syntactic characteristic measures: syntactic complexity
   1. mean T-unit length
   2. mean free modifier length
   3. mean closer length
C. Analytic score
   1. content/thinking main criteria score
   2. organization main criteria score
   3. style/tone main criteria score
   4. analytic total score
D. Holistic score

Independent Variables. The two topic types, two native language groups, and two academic major groups are used as independent variables. All these independent variables are assumed to be fixed factors in the SPF 22.2 design (two between subject treaments and one within subject treatment) and the students (blocks) variable is assumed to be a random factor. In the SPF 22.2 design, the students are nested within the native language group and the major group factors. When the topic effects on scoring method and analytic main categories are investigated, the two scoring methods and the three analytic main categories are used as dependent variables. In those case, SPF 2.22

design is used for the native language group, scoring method, and topic, and SPF 2.23 (one between subject treatment and two within subject treatments) is used for the native language group, topic, and scoring method.

A. Topic type
   1. a writing task with general information provided
   2. a writing task with narrowly defined specific information provided

B. Native language group
   1. native speakers of English group
   2. native speakers of Chinese group

C. Academic major group
   1. hard science major group
   2. social science major group

Correlational Analysis

To investigate the relationships among different measures of writing skill, the following three correlational analyses are conducted:

(a) correlations between syntactic characteristic measures and essay quality measures

(b) correlations between holistic score and analytic score

(c) correlation between writing skill measures and general verbal skill measures

The above three correlational analyses are conducted by simple correlation, stepwise multiple regression, and confirmatory factor analysis, and each analysis is separately conducted for each of the two topics.

CHAPTER IV

RESULTS

The purpose of this study was to determine whether two different topics--a topic with general information provided and a topic with narrowly defined specific information provided--will elicit a different quality of writing from a sample of upper college level students based on syntactic characteristic measures and essay quality measures. For this purpose, three major questions were raised in the previous chapter: (a) do the two different topics affect students' performance in syntactic characteristics and quality ratings in a testing situation? (b) are the topic effects consistent across different groups of writers (different native language groups and different major groups)? (c) are the relationships between different measures of writing skills remain stable across the two different topics?

In this chapter, the results of the investigation for the topic effects are reported in conjunction with the types of questions and concerns that are raised in the preceding chapters. This chapter consists of three major sections: (a) topic effects on syntactic characteristics, (b) topic effects on quality ratings, and (c) topic effects on the relationships among different measures of writing skill. Within each major section, subsections are made according to the different types of measures or scoring methods. At the end of each subsection, a brief summary and discussion are presented.

Topic Effects on Syntactic Characteristics

To investigate topic effects on syntactic characteristics of each student's essay, two types of measures--length of elaboration and syntactic complexity--were separately analyzed. For the indices of length of elaboration, three variables were used--total number of words, total number of T-units, and total number of words in free modifiers. For the indices of syntactic complexity, another three variables--mean T-unit length, percentage of words in free modifiers and percentage of words in closers--were used.

Topic Effects on Length of Elaboration

Table 2 presents means and standard deviations of the three mechanical measures of elaboration length for the total group and each of Native Language (NL) groups and Major Groups (MG). In all three variables, the "Space" topic essays have consistently higher values compared to the "Farm" topic essays through all the groups. In order to examine the topic effects on elaboration length measures across native language groups and major groups, analyses of variance based on split-plot factorial 22.2 design were performed separately on the three measures of elaboration length. ANOVA summary tables of these analyses are presented in Appendix A.

Total number of words. On the total number of words, a significant main effect is found for topic ($F = 29.8$ $df = 1,96$ $p = .001$), for NL ($F = 98.6$ $df = 1,92$ $p = .0001$), and MG ($F = 5.32$ $df = 1,92$ $p = .0233$). The results also indicate a significant interaction effect for NLxtopic ($F = 7.47$ $df = 1,92$ $p = .0075$). However, the other interaction effects are

Table 2

Mean and Standard Deviation of Elaboration Length Measures

Total Group

Variables	"Space" topic	"Farm" topic
TW	248.6 (95.5)	214.8 (82.8)
TU	14.7 (5.5)	13.1 (5.3)
WFM	60.1 (39.0)	50.3 (31.3)

Native Language Group

Variables	English		Chinese	
	"Space"	"Farm"	"Space"	"Farm"
TW	316.0 (76.8)	265.2 (75.3)	181.2 (57.0)	164.3 (54.5)
TU	16.4 (4.9)	14.1 (5.5)	13.1 (5.5)	12.1 (5.0)
WFM	84.4 (38.0)	67.0 (32.7)	35.7 (20.4)	33.6 (18.5)

Academic Major Group

variables	Hard Science		Social Science	
	"Space"	"Farm"	"Space"	"Farm"
TW	231.1 (90.3)	204.9 (81.1)	266.1 (98.3)	224.6 (84.1)
TU	14.5 (5.9)	12.7 (4.7)	14.9 (5.0)	13.6 (5.9)
WFM	52.7 (33.6)	47.5 (33.1)	67.4 (42.8)	53.0 (29.4)

Note. The variable names are as follows:

TW: Total number of Words

TU: total number of T-Units

WFM: total unmber of Words in Free Modifiers

not significant (p > .05 for MGxtopic and NLxMGxtopic). (See Appendix A for the ANOVA summary table.)

The significant topic main effect indicates that students create more semantic units for the "Space" topic than for the "Farm" topic. The other main effects show that the native-English group creates more semantic units than the native-Chinese group, and that the social-science major group creates more than the hard-science major group in this study. The significant interaction effect for NLxtopic indicates that the topic effect is stronger for the native-English group than for the native-Chinese group.

Total number of T-units. The total number of T-unit data yields significant main effects for topic (F= 9.74 df= 1,92 p= .0024) and NL group (F= 8.51 df= 1,92 p= .0044). However, MG main effect and all interaction effects (NLxtopic, MGxtopic and NLxMGxtopic) are not significant at alpha equal to .05 level. (See Appendix A for the ANOVA summary table.)

The significant topic main effect shows that students generate more independent clauses, hence more main idea units, for the "Space" topic than for the "Farm" topic. This means that when a writing task, like the "Farm" topic, provides a narrowly defined specific information, the writers probably have more constraints on the process of the text production. The nonsignificant interactions imply that the topic effects are parallel across the native-language groups and the academic-major groups. Finally, the NL main effect indicates that the native-Chinese group has more constraints on the production of main ideas than the native-English group as shown by the total number of words generated.

Total number of words in free modifiers. The total number of words in free modifiers includes all words in openers, interrupters and closers. While the total number of T-units can be regarded as the number of main idea units generated on the given writing task, the total number of words in free modifiers may be related to the number of subidea units which support the main idea units. The data on the number of free modifiers shows a significant topic main effect (F= 9.58 df= 1,92 p= .0026), NL main effect (F= 76.84 df= 1,92 p= .0001), and MG main effect (F= 5.48 df= 1,92 p= .0214). The data also yields significant interaction effects for NLxtopic (F=5.48 df= 1,92 p= .0214), and NLxMGxtopic (F= 10.53 df= 1,92 p= .0016). (See Appendix A for the ANOVA summary table.)

The significance of the topic main effect is consistent with the total number of words and the total number of T-units. This means that students use not only more semantic units and main idea units but also more subidea units for the "Space" topic which represents a general writing task than for the "Farm" topic which represents a specific writing task. The significant MG main effect indicates that the social science (including humanities) major group students use a larger number of words in free modifiers than the hard-science major group. It is probably due to their academic background which emphasizes certain types of writing tasks. The significant NLxtopic interaction effect, as in the case of total number of words, indicates that the topic effect is more stronger for the native-English group than for the native-Chinese group.

Summary and discussion. The preceding results support the notion that the amount of elaboration in writing is significantly related to the types of topics, especially with the kind of information provided in the writing assignment. In the "Space" topic, writers can draw information from their own knowledge and experience on the advantages and dis-

advantages of space exploration; whereas, in the "Farm" topic, writers have to rely heavily on the specific information given in the assignment. The results suggest that a writing task requiring writers to draw more upon their previous and existing knowledge facilitates more elaboration as measured by the total number of words, number of T- units, and number of words in free modifiers.

This results suggest that the findings of learning and memory research on reading comprehension can be applicable to writing research. Research concerned with utilizing knowledge to interpret new information has frequently employed the concept of elaboration (e.g., Anderson & Reader, 1978; Bransford, Franks, Morris & Stein, 1978; Stein & Bransford 1979 etc.). According to Stein and Bransford (1979), the concept of elaboration emphasizes the idea that learning is enhanced by relating new information to what one already knows by putting new information in the context of semantically congruous knowledge.

Effective elaboration involves the activation of information that clarifies the significance or relevance of concept relative to the events in which they occur. Several researchers report that subject-generated elaboration is more effective than experimenter-provided elaboration, although there is still considerable ambiguity regarding the effectiveness of subject-generated elaboration (Bobrow & Bower 1969; Meyer & Freedle, 1984; Slamecka & Graf 1978; Stein & Bransford, 1979).

Reading comprehension researchers also suggest that text can be presented in ways that result in more elaboration of ideas. Glover, Plake, Roberts, Zimmer, and Palmere (1981) reported that reading comprehension tasks requiring subjects to draw on previous knowledge re sulted in higher levels of text recalls than reading comprehension tasks not requiring extensive use of previous knowledge. According to their results, as the level of readers' schema and text base interaction increases, readers generated greater numbers of idea units and logical intrusions.

Glover, Plake and Zimmer (1982) investigated significant facilitative effects on readers' recall of a text through the use of higher order objectives of Bloom's (1956) taxonomy. Their results indicate that verbal material requiring more difficult decision is recalled at a significantly higher rate even when processing time is substantially controlled. The results of this study are consistent with the findings of the verbal learning and memory research cited above. The results are also similar to Benton and Blohm's (1986) finding on writing research. These researchers investigated the effects of question type (general or specific) and question position (prior to or following writing) upon measures of conceptual elaboration in terms of top-level, mid-level and base-level ideas. According to their results, questions that are specific to a particular topic do not facilitate deeper levels of processing than general questions.

In summary, with consideration of other research findings, the results of this study suggest that a writing task requiring writers to draw more heavily upon previous or existing knowledge facilitates longer elaboration as measured by total number of words, total number of T-units, and total number of words in free modifiers.

Topic Effects on Syntactic Complexity

In order to examine the topic effects on syntactic complexity of student essays, mean T-unit length and mean free modifier length (percentage of words in free modifiers and percentage of words in closers) were used as dependent variables. The mean T-unit length measure has been employed as a principal indicator of syntactic complexity in written discourse by Combs (1976), Crowhurst (1980), Gebhard (1978), Hunt (1965, 1970, 1977), O'Donnel (1976, 1977), Stewart (1978), and Witte and Davis (1980, 1982). In addition to the mean T-unit length, the percentage of words in free modifiers and

percentage of words in closers have also been used as good indicators of syntactic complexity in written discourse by Christensen (1968), Christensen and Christensen (1978), Cooper (1984), Faigley (1979), Watson (1983), and Wolk (1970). According to Watson's study, the free modifier and closer length are the most stable measures of syntactic complexity, both globally and within separate types of discourse.

Table 3 presents the mean and standard deviation of three syntactic complexity measures for the total group, each native language group and the major group.

Mean T-unit length. Mean T-unit length is the average length of T-units counted in words per T-unit. As shown in the Table 3, mean T-unit length is similar across two topics within each group. In the total group, the mean T-unit length (MTU) is higher than that of Hunt's skilled adult sample (MTU= 14.8; Hunt, 1970), but similar to Gebhard's (1978) good freshman sample (MTU = 16.3). In the native-English group, the mean T-unit length is higher than Hunt's skilled adult sample but similar to Cooper's (1984) graduate student sample (MTU= 19.4); whereas in the native-Chinese group, the length is almost the same as Hunt's skilled adult sample. The length of social-science major group is slightly higher than the length of hard-science major group.

The results of the ANOVA, based on SPF 22.2 design, indicate that the topic main effect and MG main effect are not significant (for topic main effect, F= .32 p= .9635; for MG main effect F= 3.47 p= .0657). Interaction effects of NLxtopic (F= .00, p= .9635), MGxtopic (F= .17, p= .6841), NLxMGxtopic (F= 2.49, p= .1182) are not significant. However NL main effect is significant (F= 46.23, p= .0001). (See Appendix A for the ANOVA summary table.)

Table 3

Mean and Standard Deviation of Syntactic Complexity Measures

Total Group

Variables	"Space" topic	"Farm" topic
MTU	17.4 (5.0)	17.2 (5.2)
PFM	23.1 (10.8)	23.2 (11.6)
PCL	10.2 (8.4)	8.4 (8.4)

Native Language Group

Variables	English		Chinese	
	"Space"	"Farm"	"Space"	"Farm"
MTU	19.9 (4.3)	19.7 (4.6)	14.9 (4.3)	14.6 (4.4)
PFM	26.5 (9.8)	25.7 (11.3)	19.7 (10.9)	20.7 (11.4)
PCL	13.4 (8.1)	10.7 (8.1)	7.0 (7.4)	6.0 (8.2)

Academic Major Group

Variables	Hard Science		Social Science	
	"Space"	"Farm"	"Space"	"Farm"
MTU	16.6 (5.1)	16.6 (4.7)	18.2 (4.8)	17.8 (5,6)
PFM	22.5 (11.9)	22.3 (11.6)	23.8 (9.7)	24.1 (11.6)
PCL	10.5 (8.9)	7.9 (8.7)	9.9 (8.0)	8.8 (8.2)

Note. The variable names are as follows:

MTU: Mean T-Unit length

PFM: Peicentage of words in Free Modifiers

PCL: Percentage of words in CLosers

The nonsignificant topic main effect supports the stability of the mean T-unit length across different writing tasks (e.g., general vs. specific) within individual students as a syntactic complexity measure. This result clarifies the difference between the elaboration length variables and the syntactic complexity variables. The essay length variables can be indices of the amount of idea generated, whereas the syntactic complexity variables are indices of mature style in writing. This result is consistent with Witte and Davis's finding (1982) that the mean T-unit length is stable across different writing tasks (essays developed through classification and essay developed through comparison).

The nonsignificant interaction effects of NLxtopic, MGxtopic, and NLxMGxtopic indicate that the topic effect is parallel between each of the native-language groups and academic- major groups. The significant NL main effect shows that the native-Chinese language group produces a significantly less complex style of written discourse than the native English language group. This implies that the mean T-unit length can be used as a reliable measure for differentiating between the syntactic complexity of discourses produced by different ability groups.

Percentage of words in free modifiers and closers. Table 3 shows the means and standard deviations of percentage of words in free modifiers at all positions and final position (closer) for total group and for each of the native language and academic major groups. Free modifiers are all non-restrictive modifiers (both clausal and non-clausal) including opener, interrupter and closer. The sample mean ratio of free modifiers and closers show a similar pattern for the two topics across native language and major groups. In the mean percentage of both closers and free modifiers, the native-English group shows a slightly lower percentage than Wolk's (1970) professional writer group sample (PFM = 31.0, PCL = 14.7), but a higher percentage than Wolk's college student sample (PFM = 20.9, PCL = 8.4).

The results of the ANOVA indicate that, for both free modifiers and closers, topic and MG main effects are not significant (topic main effect for free modifier, F= .01 p= .9229; topic main effect for closer, F= 2.75 p= .1006; MG main effect for free modifier, F= .89 p= .3486; MG main effect for closer, F= .02 p= .9010). The NLxtopic and MGxtopic interaction effects for both measures are not significant (NLxtopic effect for free modifier, F=.37 p= .5461; NLxtopic effect for closer, F=.57 p= .4533; MGxtopic effect for free modifier, F= .03 p= .8638; MGxtopic effect for closer, F = .46 p= .5003). Significant effects for both measures are found in NL main effect and NLxMGxtopic interaction effect (NL main effect for free modifier, F=12.73 p= .0006; NL main effect for closer, F= 22.99 p= .0001; NLxMGxtopic interaction effect for free modifier, F= 11.27 p= .0011; and for closer, F=5.12 p= .0260). (The ANOVA summary table is presented in Appendix A.)

The nonsignificant topic effect suggests that the ratio of free modifiers and the ratio of closers in upper college level students' writing is stable across two different types of writing tasks (general vs. specific) within the same discourse mode or discourse aim. The nonsignificant NLxtopic and MGxtopic interaction effects suggest that the stable trait of free modifier and closer ratio across two different types of writing tasks is parallel between different native-language groups and different major groups.

The significant NL main effect suggests that the native-English group uses a significantly higher ratio of free modifiers and closers than the native-Chinese group. This finding also supports Christensen's (1968) claim that a large percentage of free modifiers is the mark of the skillful writers. One interesting result is the significant interaction effect of NLxMGxtopic. Because of the small sample size of this study, the result is not generalizable. As shown in Figure 2, In the native-English group, the hard-science major

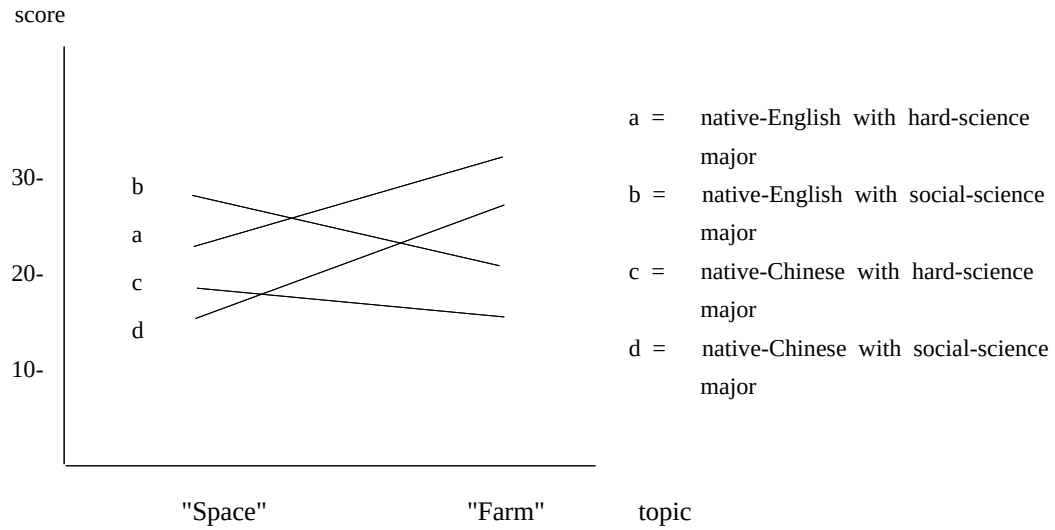


Figure 2. NLxMGxtopic interaction effect for percentage of words in free modifiers.

group students use higher percentage of free modifiers in the "Farm" topic than in "Space" topic; whereas in the native-Chinese group, the hard-science major group students use higher percentage of free modifiers in the "Space" topic than in the "Farm" topic. This result may indicate that the use of free modifiers can be affected by students' academic background, interest or experiences with the given topic.

Summary and discussion. The results clearly indicate two main points. First, syntactic complexity (measured by mean T-unit length and ratio of free modifiers) of upper college-level students' essays is stable across two different types of writing tasks (general vs. specific) within the same discourse mode or aims. Second, the stable trait of syntactic complexity across different types of writing tasks is parallel between different native language groups and between different academic major groups.

Most of the studies on syntactic complexity have been concerned with the chronological development of students' writing skill, using mean T-unit length and the ratio of free modifiers or closers as indices of syntactic complexity (maturity or fluency). These studies reported evidence that the mean T-unit length and the ratio of free modifiers or closers increased with the advances in grade level (Christensen 1968; Combs, 1976; Crowhurst, 1980; Faigly, 1979; Hunt, 1965, 1970, 1977; O'Hare, 1973; Stewart, 1978; Wolk, 1970 etc.).

However, several researchers (Crowhurst & Piche, 1979; O'Donnell, 1976; Watson, 1983) raised a question about the stability of syntactic complexity measures across different writing situations because the nature of the writing task may influence the mean T-unit length or the ratio of free modifiers as much as the individual student's grade level. As Witte and Davis (1982) pointed out, the issue of individual and group stability of mean T-unit length or ratio of free modifier across different writing situations is important because instability of these measures of syntactic complexity across different writing situations may lead the findings of preceding syntactic complexity studies to be unreliable and useless for instructional decisions in writing programs or evaluation.

In keeping with the research question, it has been reported that syntactic complexity in written composition is significantly affected by different discourse modes and different intended audience variables (Crowhurst and Piche, 1979; Watson, 1983). In spite of the claim made for discourse mode effects on syntactic complexity measures, one unsettled question remains Are the syntactic complexity measures unstable across different types of writing tasks within the same discourse mode (description, narration, exposition, or argument)? The present study has shed light on this question. According to the results

of this study, the syntactic complexity measures are stable across different types of writing tasks within the same discourse mode.

Topic Effects on Analytic and Holistic Scores

In the preceding section, the results of the investigation of topic effects on syntactic characteristics were presented and discussed in relation to the essay length and the syntactic complexity based on mechanical counting. In this section, the results of the investigation for topic effects on the quality ratings of student essays are presented and discussed in regard to the analytical ratings and the holistic ratings based on raters' subjective judgment. In addition, the results of the investigation for the reading consistency (reliability) for the two topics, and the topic and scoring method interaction effects are presented and discussed.

The results are presented in the two subsections: (a) topic effects on the analytic scores, and (b) topic effects on the holistic scores. For the topic effects on the analytic scores, the three analytic main criteria scores and the analytic total scores are separately analyzed in terms of topic main effect, writer group main effect, and topic writer group interaction effect. The topic effects and other interaction effects are analyzed using an analysis of variance technique based on a split-plot factorial design, as in the case of topic effects on syntactic characteristics. In this design, topic, native-language group, major group, scoring method, and rater group are assumed to be fixed factors, and students (blocks) are assumed to be a random factor. The ANOVA summary tables are presented in Appendix A.

Topic Effects on Analytic Score

Analytic scores of this study were generated by two rater groups using Purves's analytic scoring method (1985). Each rater group scored two different essays from each student ("Space" topic essay and "Farm" topic essay). Within each rater group, one rater scored odd numbered student essays, and the other rater scored even numbered student essays. The analytic scoring method included three main criteria and thirteen subcriteria. The analytic scores of each essay were originally generated on the basis of thirteen subcriteria. However this study deals mainly with the three main criteria and the total scores. In this study, the content/thinking main criteria scores were based on mean scores of seven subcriteria scores, and the organization main criteria and style/tone main criteria scores were based on mean scores of each three subcriteria scores. Total scores were based on the sum of the three main criteria scores. Therefore, each main and subcriterion scores ranges from 1 to 5 point, and the total analytic scores range from 3 to 15 points. In the process of interpreting the results, the rating schemes and the standards of high quality essays for each main and subcriteria are briefly explained. These explanations are based on Purves' (1985) scoring framework and Soter's (1985) report on rating procedure.

Content and Thinking main criterion. The ratings of content/thinking main criterion focused on what was written and the way it reflected the student's manipulation of the topic. Table 4 presents the means and standard deviations of content/thinking main criterion scores for each group, and the means and standard deviations of content/thinking subcriteria scores for total group.

Table 4

Mean and Standard Deviation of Content/Thinking Main Criterion and Subcriteria Scores

Content/Thinking Main Criterion Scores

Group	"Space" topic	"Farm" topic
Total group	2.66 (.73)	2.54 (.71)
Native-English group	3.20 (.49)	3.07 (.50)
Native-Chinese group	2.11 (.48)	2.02 (.45)
Hard-science group	2.63 (.71)	2.46 (.68)
Social-science group	2.68 (.76)	2.63 (.73)

Content/Thinking Subcriteria Scores for Total Group

Subcriteria	"Space" topic	"Farm" topic
Adequacy of information	2.73 (.75)	2.71 (.91)
Richness of information	2.57 (.98)	2.42 (.78)
Relationships	2.79 (.81)	2.77 (.79)
Inferences	2.84 (.82)	2.89 (.80)
Synthesis	2.52 (.79)	2.56 (.80)
Evaluation	2.76 (.74)	2.42 (.76)
Alternatives	2.40 (.73)	2.04 (.64)

62

Table 4 shows that the mean scores of the content/thinking main criterion of the "Space" topic are consistently higher than those of the "Farm" topic in the total group and each of the native-language groups and major groups. In five of the seven subcriteria, the "Space" topic mean scores are higher than the "Farm" topic mean scores; however, in the inferences and synthesis subcriteria, the "Farm" topic mean scores are slightly higher than the "Space" topic mean scores.

The results of the ANOVA show significant topic main effect (F=4.26 df= 1,92 p= .0419) and NL main effect (F= 176.15 df= 1,92 p= .0001) for the content/thinking main criterion scores. However, MG main effect and all interaction effects (NLxtopic, MGxtopic, NLxMGxtopic) are not significant at alpha equal to .05 level. (See Appendix A for the ANOVA summary table.)

The significant topic main effect indicates that students' essays on the "Space" topic (a writing task with general information provided) are rated more highly than essays written on the 'Farm" topic (a writing task with specific information provided) in regard to the content/thinking main criterion. The nonsignificant interaction effects indicate that the topic effect is parallel across each of the native language and major groups.

Further analyses were conducted to investigate which subcriteria scores of the content/thinking dimension are significantly affected by different types of writing tasks. According to the results, the significant topic effects are found in three subcriteria: richness of information (F=4.15 df= 1,92 p= .0445), evaluation (F=21.8df=1,92p= .0001), and alternatives (F=23.64df=1,92p=.0001). These significant topic effects can be explained by examining the relationship between the different types of writing tasks and the scoring standards for the subcriteria.

The rating of the richness of information subcriterion was based on the use of additional information drawn from a variety of sources such as reading or general knowledge, or the amount of richness thought to be desirable in relation to the topic. For the "Space" topic, high scores would be given to essays which provide a detailed context for the argument, a detailed definition of space exploration, or past and present public attitudes toward space exploration. Because the writing assignment did not provide these kinds of detailed information, all this information had to be drawn from the students' personal knowledge or experience. For the "Farm" topic, high scores would be given to essays which provided detailed referpnces to the kind of graphic representation used, such as the actual increase and decrease in the four decades, and the implication for production and employment, and population shifts. However, it may be difficult for a student to provide this kind of information for the "Farm" topic because the task of this topic requires students to focus on the specific information provided in the graphs. The task itself may limit the free use of students' background knowledge. As in the case of length of elaboration measures, the significant topic effect for this subcategory indicates that when a writing task with general information is given to students, students generate- higher quality essays in terms of richness of information than when a writing task with specified information was given.

The rating of the evaluation subcriterion was based on the degree to which students appeared to make judgments as to the relative merits of particular relationships, inferences, or synthesis, and the degree to which applicable criteria were used. In the "Space" topic, students were required to provide two opposing views and to adopt one of them, giving reasons for the position adopted. Therefore, the task of the "Space" topic requires students explicitly to make evaluations and judgments about the relative merits of their relationship and inferences. In the "Farm" topic, students could evaluate the content,

relationships, and inferences. They could evaluate the correctness of relationships in the graphs, or they could evaluate their conclusion by giving counter arguments or judgments backed up with examples or explanation. However, the "Farm" topic task does not explicitly require students to provide such evaluation in their essays. Therefore, the significant topic effect for the evaluation criterion indicates that the quality ratings of students' essays, in terms of evaluation, can be affected by different task requirements (broadly defined task vs. narrowly defined task).

The rating of the alternatives criterion was based on the degree to which the writers appeared to admit the possibility of alternative arguments or interpretations, and either accept them as admissable or oppose them. In the "Space" topic, students were required to discuss both positions (pros and cons) of space exploration; in this way, they were to show awareness of the limitations and possible alternative explanations in their essays. In the "Farm" topic, students were to show alternative interpretations of the graphs which might have indicated several possible outcomes or different things. However, it may be very difficult to interpret the graphs in several different ways since the interpretation of the graphs seems to be straightforward because of the narrowly defined writing task and specific information provided. For example, an increase in farm size causes the introduction of bigger farm machinary and a simultaneous decrease in farm population. Therefore, in the "Farm" topic, the possible alternatives would be limited and not be considered an important task by the students. The result indicates that a topic with a more general (broadly defined) task may obligate students to produce more alternative ways of thinking.

The nonsignificant topic effects on subcriteria such as adequacy of information($F = .04$ $p = .8358$), relationship($F=.10$ $p=.7559$), inferences($F=.49$ $p=.4847$), and

synthesis (F= .29 p= .5884) may imply that these subcriteria are not sensitive to these writing tasks (in terms of general vs. specific task).

The adequacy of information criterion was rated on the basis of thed egree to which all of the relevant information from the assignment was expressed in the essay. For the "Space" topic, a highly rated essay would provide some background information on space exploration, the extent of space exploration over time, and the benefits and limitations of space exploration so far. All of this information should be drawn from a student's knowledge about space exploration. For the "Farm" topic, a highly rated essay would provide information by referring to the changes in farm size, farm population, and the distribution of farms for each of the four decades given in the graphs. Most of this information can be drawn from the assignment. Therefore, the nonsignificant topic effect indicates that the quality ratings regarding adequacy of information is not significantly affected by the different information sources such as students' knowledge or detailed information given in the writing task.

The rating of the relationship criterion was based on the degree of connections which students made between or among the various items of information and the validity and complexity of the relationships. For the "Space" topic, a highly rated essay would provide a clear relationship between space exploration and current desirability of such exploration, and the relationship between current problems and long-term benefits. All these relationships should be made on the basis of inferences drawn from students' knowledge. In the "Farm" topic, a highly rated essay would relate the information from one graph to the others, such as farm size, farm population, and number of farms. It would also make clear the links between the interpretation of one graph and the others. All of these relationships should be made on information provided in the assignment.

Therefore, as in the case of adequacy of information subcriterion, the nonsignificant topic effect indicates that the quality ratings regarding the relationship is not significantly affected by different information sources such as students' knowledge or detailed information given in the writing task.

The inferences subcriterion scores were generated on the basis of the kind of interpretation that students made above and beyond the information provided in the assignment. The "Space" topic required students to make inferences about the advantages and disadvantages of space exploration. The "Farm" topic required students to make inferences about the major outcomes of changes in population, size of farms, and number of farms over a forty-year period as indicated by the graphs. Because both topics required students to make inferences beyond the information given in the assignment, different types of tasks would not affect the scores of this criterion.

The synthesis criterion scores were generated on the basis of the degree to which students appeared to draw together the information, relationships, and inferences into a single complex generalization. For the "Space" topic, a highly rated essays would discuss the advantages and disadvantages of space exploration and make logical connections between these two positions rather than simply presenting them both. In the "Farm" topic, a highly rated essay would provide an integrated discussion of the changes in farm size, population, and number of farms; and would show the implications and consequences of such changes. The result indicates that the scores of this criterion are not significantly affected by the different types of tasks. This may be due to the fact that the synthesis criterion requires highly complicated cognitive skills beyond the information provided in the writing assignments.

In summary, the topic effect on the analytic content/thinking main criterion is significant, indicating that the "Space" topic is rated more highly than the "Farm" topic. This result is derived from the subcriteria--richness of information, evaluation, and alternatives--used in the scoring of the content/thinking dimension. A careful examination of the subcriteria shows that these subcriteria are more well suited for the "Space" topic (with general information) than for the "Farm" topic (with narrowly defined specific information).

Organization main criterion. The rating of the organization main criterion concerned the structures of written essays both as a whole essay and its various parts. Table 5 shows the means and standard deviations of the content/thinking main criterion scores for each group, and the mean and standard deviation of organization subcriteria scores for the total group. Table 5 shows that the mean scores of the organization main criterion of the "Space" topic are slightly higher than those of the "Farm" topic for the total group and for each of the native-language groups and major groups. In the three sub- criteria, the "Space" topic mean scores are also slightly higher than the "Farm" topic mean scores.

The result of the ANOVA shows a significant NL main effect ($F= 117.24$ df= 1,92 $p= .0001$) for the organization main criteria scores. However, topic main effect ($F= .78$ df= 1,92 $p= .3790$), MG main effect ($F=2.41$ df= 1,92 **p=** .1244), and all interaction effects (NLxtopic, MGxtopic, NLxMGxtopic) are not significant at alpha equal to .05 level. The nonsignificant topic main effect indicates that the ratings of the students' essays on the "Space" topic (a writing task with general information provided) and the ratings of the students' essays on the "Farm" topic (a writing task with specific information provided) are not significantly different for the organization main criterion. The

Table 5

Mean and Standard Deviation of Organization Main Criterion and Subcriteria Scores

Organization Main Criterion Scores

group	"Space" topic	"Farm" topic
Total group	2.77 (.76)	2.72 (.77)
Native-English group (NL1)	3.33 (.52)	3.31 (.53)
Native-Chinese group (NL2)	2.22 (.51)	2.14 (.46)
Hard-science group (MG1)	2.72 (.71)	2.65 (.74)
Social-science group (MG2)	2.83 (.81)	2.80 (.80)

Organization Subcriteria Scores for Total Group

Subcriteria	"Space" topic	"Farm" topic
Framing	2.92 (.85)	2.80 (.81)
Grouping	2.69 (.77)	2.67 (.82)
Unity	2.72 (.87)	2.71 (.78)

nonsignificant interaction effects indicate that the topic effect for the organization scores is parallel across each of the native-language and major groups. (See Appendix A for the ANOVA summary table.)

Further analyses were conducted to investigate whether all of the subcriteria scores of the organization dimension were not significantly affected by different types of writing tasks. According to the results, topic effects are not significant in all three subcriteria: framing($F=3.23 df=1,92 p=.0754$), grouping($F=.08 df=1,92 p=.7797$), and unity($F=.03 df=1,92 p=.8735$). These nonsignificant topic effects may be explained by examining the relationship between the characteristics of essay topics and the scoring scheme for the subcriteria.

The rating of the framing criterion was based on the degree to which students presented a context so that the essay had a beginning, middle, and end. The highly rated essays on the "Space" topic could show that two positions are possible and arguable, and the framing of the context could be made by showing the advantages and disadvantages at some point in the paragraph. Highly rated essays on the "Farm" topic could begin with a general comment about the important changes in farming patterns, and could provide main supporting points and supply a concluding statement within the body of essay. The framing of the "Farm" topic is much easier than the framing of the "Space" topic, but the topic effect is not significant for the framing subcriterion. This result may be due to the fact that the assignment of the "Farm" topic includes the contexts for framing in the forms of graphs and directions, so that students would probably pay little attention to framing of the "Farm" essays.

The rating of the grouping subcriterion was based on the degree to which students combined the information, relationships, and inferences. In the "Space" topic, highly rated essays would discuss the advantages and disadvantages of space exploration, or they would provide counter arguments. In the "Farm" topic, highly rated essays would combine a discussion of the three graphs by interpreting the points of significant changes. So, both topics require students to organize their main ideas by grouping the information provided by their long term memory or the assignment itself. Therefore, the quality of essays in the grouping criterion would not be significantly changed as a result of different types of information given in the assignments.

The rating of the unity subcriterion was based on the degree to which students provided cohesive information, showed relationships, made inferences, and excluded unnecessary information. For the "Space" topic, highly rated essays would provide a

cohesive context for the topic within which the advantages and disadvantages of space exploration would be viewed in the general context of human efforts. For the "Farm" topic, highly rated essays would provide explicit connections between the explanations of the graphs and their relationships with each other and inferences drawn from them. As in the case of grouping subcriterion, both topics require students to make clear connections between the main ideas of the writing beyond detailed information provided from their long-term memory or the assignment itself Different types of discourse modes, such as description, narration, and exposition etc., may require students to use different types of organizational strategy, but the writing tasks of this study required the same discourse mode. Therefore, as the results indicate, the quality of the students' essays regarding the unity criterion is not significantly changed by the different types of information provided in the assignments.

In summary, the topic effect on the analytic organization main criterion is not significant, indicating that the essay quality in the dimension of the organization is not changed significantly from one topic (general information) to the other topic (specific information). An examination of the scoring standard in the organization subcriteria shows that the quality in the organization subcriteria may not be significantly changed by the different types of information provided in the assignment. This result further implies that the organizational strategies used in the production of the essays are not sensitive to the different type of information provided in the assignments.

Style and Tone main criterion. The rating of the style/tone main criterion concerned the manner in which the essay was presented, particularly the degree to which the manner matched the conventions of academic discourse of English in the United States. The scores of this main criterion consist of the mean of the three subcriteria: objectivity,

tentativeness, and metalanguage. Table 6 presents the mean and standard deviation of the style and tone main criterion scores for each group, and the mean and standard deviation of the style and tone subcriteria scores for total group.

Table 6 shows that mean scores of the style/tone main criterion for the "Space" topic are slightly lower than those for the "Farm" topic in the total group and each of the native-language group and major group. In the tentativeness and metalanguage subcriteria, the "Space" topic mean scores are higher than the "Farm" topic mean scores. However, in the objectivity subcriterion, the "Space" topic mean scores are lower than the "Farm" topic mean scores.

The result of the ANOVA shows a significant NL main effect (F= 199.49 df= 1,92 p= .0001) for the style and tone main criteria scores. However, topic main effect (F= 1.50 df= 1,92 p= .2230), MG main effect (F= 1.68 df= 1,92 p= .1982), and all interaction effects (NLxtopic, MGxtopic, NLxMGxtopic) are not significant at alpha equal to .05 level. The nonsignificant topic main effect indicates that the ratings of students' essays on the "Space" topic (a writing task with general information provided) and students' essays on the "Farm" topic (a writing task with specific information provided) are not significantly different in the style/ tone main criteria. The nonsignificant interaction effects indicate that the topic effect for the style/tone score is parallel across each of the native language and major groups.

Further analyses were conducted to investigate whether all of the subcriteria scores of the style/tone dimension were not significantly affected by different types of writing tasks. According to the results, the topic effect is significant in the objectivity criterion (F= 31.44 df= 1,92 p= .0001). However, topic effects were not significant in the other

Table 6

Mean and Standard Deviation of Style and Tone Main Criterion and Subcriteria Scores

Style/Tone Main Criterion Scores

Group	"Space" topic	"Farm" topic
Total group	2.54 (.77)	2.60 (.67)
Native-English group	3.12 (.60)	3.14 (.42)
Native-Chinese group	1.96 (.40)	2.07 (.39)
Hard-science group	2.51 (.78)	2.53 (.65)
Social-science group	2.57 (.78)	2.68 (.70)

Style/Tone Subcriteria Scores for Total Group

Subcriteria	"Space" topic	"Farm" topic
Objectivity	2.66 (.82)	3.05 (.80)
Tentativeness	2.77 (.76)	2.32 (.76)
Metalanguage	2.54 (.77)	2.44 (.74)

two subcriteria: tentativeness (F=1.31 df= 1,92 p= .2550), metalanguage (F = 2.42 df= 1,92 p= .1233). The significant topic effect for the objectivity subcriterion, and nonsignificant topic effects for the tentativeness and metalanguage criteria may be explained by examining the relationship between the characteristics of essay topics and the scoring scheme for the subcriteria.

The rating of the objectivity criterion was based on the use of impersonal and detached language as opposed to personal and emotional language. For the"Space" topic, the task requires students to keep some degree of objectivity in providing a position with reasons for taking that position. However, the task requires students to take a position and defend it, calling some degree of subjectivity. In the "Farm" topic, the task requires that students make interpretations which should be supported by reference to the

graphs. Hence, objectivity is necessary for the "Farm" topic. It is expected that the scores for the interpretive task (with specific information) would be higher than the scores for the argumentative task (with general information) in the objectivity criterion. The results support this expectation.

The rating of the tentativeness criterion was based on the use of semantic hedgers and qualifiers that are often considered appropriate in academic writing. In the "Space" topic, highly rated essays were to evidence a high degree of tentativeness through writer's detaching themselves from either position and making no firm judgment about the relative merits or demerits of space exploration. In the "Farm" topic, highly rated essays would show that the interpretations were supported by facts drawn from the graphs. The writer would be aware of possible limitations in interpretation and would avoid absolute terms. Because the task of the "Farm" topic focuses on interpretation based on specific information provided in the assignment, and because it requires objectivity, a consistently higher degree of tentativeness would be expected in the "Space" topic. However, the result indicates that the rating of this subcriterion is not significantly affected by different types of writing tasks.

The rating of the metalanguage subcriterion was based on the use of markers to indicate the relationship between propositions and paragraphs. For both the "Space" topic and the "Farm" topic, highly rated essays used markers (such as however, therefore, etc.) to show the logical relationships between argumentation or interpretation. Therefore, the rating of this subcriterion would not be affected by different types of writing tasks within the same discourse mode. The results support this expectation.

Taken as a whole, according to the scoring standard used in the analytic style/tone subcriteria, the essay quality in the dimension of the style/tone may not be significantly changed by the different types of information (general vs. specific).

Total Analytic scores. The total analytic scores were generated based on the sum of three main criteria scores. Table 7 presents the mean and standard deviation of the total analytic scores for each group. Table 7 shows that the means of the "Space" topic total analytic scores are slightly higher than those of the "Farm" topic in the total group and for each of the native-language group and major group. In the major group, the mean of the social-science group is higher than the mean score of the hard-science group as in the case of each main criterion score.

The result of the ANOVA shows a significant NL main effect (F= 203.87 df= 1,92 p= .0001) for the analytic total scores. However, the topic main effect (F= .50 df= 1,92 p= .4824), the MG main effect (F= 2.25 df= 1,92 p= .1367), and all interaction effects (NLxtopic, MGxtopic, NLxMGxtopic) are not significant at alpha equal to .05 level. The nonsignificant topic effect indicates that when the quality rating is conducted by the analytical scoring method, the quality of students' essays is not significantly affected by different types of writing tasks in regard to the types of information provided in the assignments. The results further suggest that when students' performance on writing is decided on the basis of the total analytic score in a testing situation, the two topics may provide redundant information, so only one topic is sufficient for the purpose. The nonsignificant interaction effects indicate that the topic effect for the total analytic score was parallel across each of the native-language and major groups. This nonsignificant interaction implies that the stability of analytic total scores can be applicable to different

Table 7

Mean and Standard Deviation of Analytic Total Scores

Group	"Space" topic	"Farm" topic
Total group	7.97 (2.20)	7.87 (2.08)
Native-English group	9.66 (1.51)	9.52 (1.38)
Native-Chinese group	6.28 (1.29)	6.23 (1.20)
Hard-science group	7.86 (2.14)	7.63 (2.00)
Social-science group	8.08 (2.27)	8.11 (2.15)

native-language groups or major groups. (See Appendix A for the ANOVA summary table.)

Further analysis was conducted to investigate whether the topic effect on the three main criteria was significantly different from each others. For this analysis, a split-plot factoprial 2.23 (two native language groups, two topics, and three main criteria) design was used. (See Appendix A for the ANOVA summary table.) The results indicate a significant main criteria effect (F=36.49 df= 2,188 p= .0001), and topicxcriteria interaction effect (F=6.82 df= 2,188 p= .0014). In both topics, the organization criterion scores are the highest; however, for the "Space" topic,the content/thinking scores are higher than the style/tone scores. For the "Farm" topic, the style/tone scores are higher than the content/thinking scores (see Figure 3).

The native-language group main effect is significant in the analytic total scores as well as the three main criteria scores. This result indicates that the native-English group students have a significantly higher ability in writing skills than the native Chinese group students. However, one interesting question remains: when the analytic score is adjusted

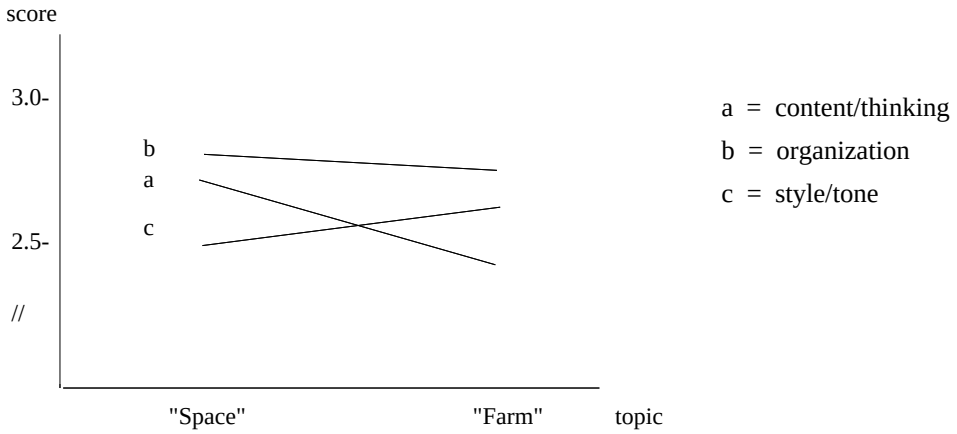


Figure 3. TopicxCriteria interaction effect in the analytic main criteria scores.

for the general verbal skill score, is the group difference still significant? The answer for this question may provide information about the nature of writing skill. For this question, an analysis of covariance for randomized block design was conducted, using the GRE verbal score as a covariate, and using the total scores of the two topics as a dependent variable. According to the result, the group difference is still significant (F= 18.03 df= 1,46 p= .0001). (See Appendix A for the ANCOVA summary table.) This result implies that the group difference in writing skill (measured by an essay test) is stronger than the group difference in the general verbal skill (measured by an objective test).

Reading consistency on the analytic score of two topics. Several researchers suggest that the ratings of student essays would vary for different types of writing tasks. To examine the topic effect on readers' ratings, this study conducted two types of analyses : usual F test based on SPF 2.22 design (2 native language groups, two rater groups, and two topics), and reliability estimation *(see* Appendix A).

According to the results of the F test, the Rater Group (RG) main effect (F = 1.90 df= 1,94 p= .1718), the RGxtopic interaction effect (F= .85 df= 1,94 p= .3584), and the NLxRGxtopic interaction effect (F= .07 df= 1,94 p= .7970) are not significant. The RG main effect indicates that the two rater groups show no significant differences in their ratings of student essays, and the RGxtopic interaction effect shows that two rater groups are consistent in their ratings of the two different topics. However, the result of the ANOVA shows a significant NLxRG interaction effect (F= 11.78,_ df= 1,94 p= .0009). In the native-English group, the first rater group's mean score is higher than the second rater group's mean score, but in the native-Chinese group, the first rater group's mean score is lower than that of second rater group. This implies that although different raters' ratings are stable across different types of topics, the ratings can be affected by writer groups with a different cultural background (see Figure 4).

Topics do not significantly affect the reading reliabilities on the total analytic scores. (for "Space" topic, rel= .80 ; for "Farm" topic, ref= .81). This means that reading performance across different topics is stable, so that the rating of the two rater groups is equally reliable for both topics (see Table 8). However, reading reliabilities of the main criteria for the two topics are not consistent. For the "Space" topic, content/thinking criteria reliability is the highest (ref = .79); whereas, for the "Farm" topic, style/tone criteria reliability is the highest (rel= .83). This implies that reading reliability of analytic main criteria scores can be affected by different type of writing tasks. An implication of this result is that the reading reliability is generally better for the scoring criterion judged to be of major importance to the essay than for the scoring criterion judged to be of minor importance (Linn, Klein, & Hart, 1972).

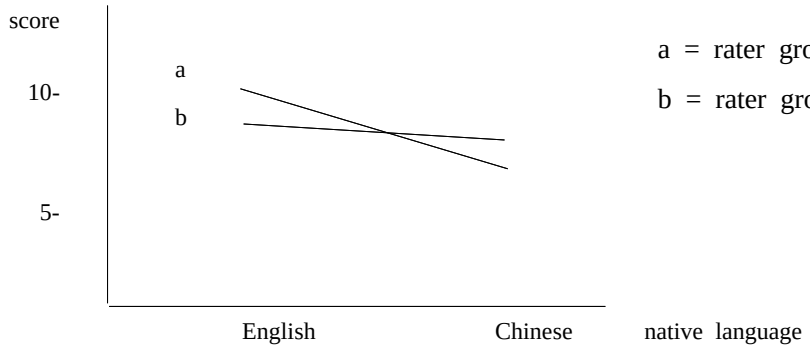


Figure 4. Rater GroupxNL interaction effect in the total analytic score.

Table 8 shows that reading reliabilities can be increased by using two more raters (from .80 to .91). Although four raters per essay might yield a higher reading reliability, it may be too time consuming and costly to have each essay read by more than two different readers in many testing situations. The rater reliabilities of this study are higher than those of other researchers' reports. According to Breland, reading reliability of the two raters for the analytic total scores was .78, and the range of reading reliabilities for subscale was from .69 to .71. The reading reliabilities of the analytic scores in this study are also higher than other researchers' reading reliability estimates for holistic scores in the case of two readings for one task (e.g., Coffman, 1971b, rel= .56 ; Breland, 1983, rel= .76).

However, the reading reliability of the analytic scores in this study is lower than the reading reliability of the holistic scores in this study (for the "Space" topic, rel= .81, and for the "Farm" topic, rel= .87 in two readings for the holistic scores). This results do not supports other researchers' findings that the analytic scoring method

Table 8

Reliabilities of Analytic Scores

Reading Reliabilities for Each Topic

Category/ topic	"Space" (2 raters)	"Farm" (2 raters)	"Space" (4 raters)
Content/ thinking	.79	.75	.90
Organization	.72	.71	.88
Style/tone	.73	.82	.88
Analytic total	.80	.81	.91

Reliabilities of Two Topics with Two Readings

Score reliability	.88
Reading reliability	.89
Coefficient of generalizability	.88

produces a higher reading reliability than those using an holistic scoring method (Bauer, 1982; Breland, 1983).

Table 8 also presents the score reliability and the rater reliability for two reading with two tasks. These reliabilities were generated based on an analysis of variance using the model referenced by Godshalk et al. (1966). Reliability estimation procedures are presented in Appendix A. The score reliability for the topics is .88, and the reading reliability for the two topics is .89. The score reliability can be interpreted as a true correlation between the scores of the same students if they were to write on two similar topics and have them scored by two readers. As already reported in the analysis of topic effect on total analytic scores, the score reliability of this sample indicates that the sample students' performance does not vary significantly from topic to topic. The nonsignificant interaction between students and topics ($F=1.06$ $df= 95,95$ $p= .3875$) also indicates that students perform similarly on both topics. The coefficient of

generalizability for the data is similar to the score reliability. Because the topics and rater groups are uniform for all students, observed variance of the data is equal to the sum of four variance components: students variance, studentxtopic variance, studentxrater variance, and studentxraterxtopic variance. The coefficient of generalizability was estimated by dividing the student variance by the observed variance (Thorndike, 1982). The largest component of the error variance was studentxraterxtopic variance, and the error variances of studentxtopic and studentxreading were negligeable.

Summary and discussion. The results of the investigation for the topic effects on the analytic score indicate the following major points:

1. Of the three analytic main criteria scores, only the content/thinking main criterion score was significantly affected by the two different types of topics. Among the seven subcriteria of the content/thinking main criteria, richness of information, evaluation, and alternatives subcriteria scores are significantly higher for the "Space" topic than for the "Farm" topic. All of these subcriteria required (in order) the use of additional information drawn from a variety of sources in relation to the topic, the use of judgments about the relative merits of particular relationships, inferences or synthesis, and the use of alternative arguments or interpretations. Therefore, the significant topic effects on the content/thinking main criterion and those subcriteria suggest that the 'general' topic (a topic with general information provided) may obligate students to produce more extensive information, more explicit judgment, and more alternative views of thinking than does the 'specific' topic (a topic with narrowly defined specific information provided).

2. The organization main criterion and the style/tone main criterion scores are not significantly affected by the two different types of topics. The results indicate that different types of information (general vs. specific) provided in the topics do not significantly affect students' performance in regard to the structure of written essay as a whole

and its various parts, or in regard to the manner that matches the conventions of academic discourse in English. This result implies that the two main criteria are not sensitive to the kind of information provided in the writing assignment. In addition, the result may indicates that the content/thinking category focuses on declarative knowledge for the writing, and the organization and style/tone categories focus on procedural knowledge for the writing. The declarative knowledge on the content of the essay may be sensitive to the types of information given in the assignment, but the procedural knowledge on the organization and style of the essay may not be sensitive to the types of information given, in the assignment.

3. The comparison of the three main criteria scores shows a significant interaction effect of topic and main criteria. This suggests that the student performance on each main criterion fluctuates from topic to topic.

4. The topic effect on the total analytic score was not significant. This indicates that when students performance is evaluated on the basis of the total analytic score, the two types of topic provide redundant information; hence, one topic is enough for that purpose.

5. All of the interaction effects of NLxtopic, MGxtopic, and NLxMGxtopic were not significant, indicating that the topic effects of each main criterion score and total analytic score were parallel across different native-language groups and across different major groups.

6. The inspection of reading (rating) consistency suggests that the two rater groups consistently rated each student's essay within and between the topics. The analysis of variance results show that the two rater groups' ratings are not significantly different from each other within and between the topics. In addition, the rater reliability and score reliability for the sample of this study was comparatively higher than those of other studies. This may validate the use of this sample for purposes of investigating the topic

effects on student writing. However, the reading reliabilities of the analytic main criteria scores are affected by the different types of writing tasks.

The above results are closely related to the finding of the preceding section, which suggests that a writing task requiring writers to draw more heavily upon their own knowledge (or long-term memory) facilitates longer elaboration as measured by total number of words, T-units, and free modifiers. Although several researchers report that students' performance fluctuated from one essay to another, the above finding differ from the previous researchers' finding, which have usually been based on holistic scores. The result of this study may provide detailed information about the specific mechanisms which cause fluctuation of writing quality from one topic to the other.

Topic Effects on Holistic Scores

The holistic scores of this study were originally generated on a six-point scale by two raters at the Educational Testing Service (ETS). The two raters' average scores were used in this study to examine the topic effects on holistic scores. The holistic scoring method, developed by ETS, requires no detailed discussion of subcriteria and no summing of scores given to separate subscales. The rater independently scored students essays by judging where the paper fits within the range of essays written for a given assignment. In cases of extreme disagreement between raters, a third reader are used to resolve the difference (Breland & Griswold, 1981; Cooper, 1977). The rater is usually trained to read the essay quickly and to score it as a whole, without considering the several dimensions of writing skills. Therefore, the holistic scores represents the rater's opinion of the overall quality of the essay (Freedman & Calfee, 1983). The holistic scoring (used in this study) dealt with mechanical eroors; whereas, the analytic scoring

(used in this study) did not consider mechanical errors. In this section, the results of the investigation for topic effects on holistic scores and for topic effects on different scoring methods are reported and discussed.

Holistic scores. Table 9 presents the mean and standard deviation of the holistic scores for each group and for the total group. The means of the "Space" topic holistic scores are lower than those of the "Farm" topic in the total group and for each of the native-language group and major group. It is clearly different from the analytic score. In the major group, the mean of the social-science group is higher than the mean score of the hard-science group.

The results of the ANOVA show a significant topic main effect (F=5.66 df= 1,92 p= .0194), NL main effect (F= 284.70 df= 1,92 p= .0001), and MGxtopic interaction effect (F = 5.07 df= 1,92 p= .0267) for the holistic scores. However, MG main effect (F= 3.26 df= 1,92 p= .0741), NLxtopic interaction effect (F= .93 df= 1,92 p= .3372), and NLxMGxtopic interaction effect (F= .69 df= 1,92 p= .4351) are not significant. The significant topic effect indicates that when the quality rating is conducted by a holistic scoring method, the quality of student essays can be significantly affected by the different topics in regard to the types of information provided in the assignments. The significant MGxtopic interaction effect indicates that the hard-science major group students receive a significantly higher holistic scores on the "Farm" topic than on the "Space" topic; while the holistic scores of the social-science major group students are not significantly different across different types of topics. They receive s lightly higher scores for the "Farm" topic than they did for the "Space" topic. (See Appendix A for the ANOVA summary table.)

Table 9

Mean and Standard Deviation of Holistic Scores

Group	"Space" topic	"Farm" topic
Total group	4.27 (1.55)	4.46 (1.55)
Native-English group	5.52 (.76)	5.79 (.50)
Native-Chinese group	3.02 (1.05)	3.13 (1.01)
Hard-science group	4.04 (1.54)	4.41 (1.40)
Social-science group	4.50 (1.55)	4.51 (1.71)

Holistic score and Analytic score. The above results indicate clear differences between topic effects on the holistic score and topic effects on the analytic score. In the analytic score, the "Farm" topic scores are generally lower than the "Space" topic scores, however, in the holistic score, the "Farm" topic scores are generally higher than the "Space" topic score. To investigate this phenomenon, the analytic score was transformed into a six point scale like the holistic score.

Table 10 presents the mean score of the transformed analytic scores and holistic scores for the two topics, which show that the holistic score is comparatively higher than the transformed analytic score. The "Space" topic analytic score is higher than the "Farm" topic holistic score in each group; whereas, the "Space" topic holistic score is lower than the "Farm" topic holistic score in each group.

To examine the topic effect on the scoring methods, an analysis of variance was conducted based on split plot factorial 2.22 design (two native-language groups, two topics, and two scoring methods) (see Appendix A). The results show a significant scoring method (SM) main effect ($F= 367.29$ $df= 1,94$ $p= .0001$), topicxSM interaction

Table 10

Comparison of Holistic Score and Transformed Analytic Score

	Analytic score		Holistic score	
	"Space"	"Farm"	"Space"	"Farm"
Total group	3.19	3.11	4.27	4.46
Native English	3.86	3.77	5.52	5.79
Native Chinese	2.52	2.47	3.02	3.14

effect (F= 8.09 df= 1,94 p= .0055), and NLxSM interaction effect (F= 101.91 df= 1,94 p= .0001).

The significant topicxSM interaction suggests that the "Space" topic is rated highly by the analytic scoring method, whereas the "Farm" topic is rated highly by the holistic scoring method. The possible reasons for this difference will be further investigated in the next section. The significant NLxSM interaction effect indicates that the scoring method effect is more significant for the native-English group than for the native- Chinese group.

Summary and discussion. The results of the investigation for the topic effects on the holistic score provide the following four major points:

1. The holistic scores are significantly affected by the different types of topics. Students receive significantly higher scores for the "Farm" topic than they did for the "Space" topic, and the significant topic effect is clearer in the native-English language group.

2. The pattern of topic effect on the holistic scores contrasts with the topic effect on the analytic score. The analytic score of the "Space" topic is generally higher than the "Farm" topic score; however, the holistic score for the "Space" topic is significantly lower than the "Farm" topic score.

3. The two major groups' holistic scores are significantly affected by the two different topic types. The holistic score difference between the two topics is significant in the hard-science group; whereas, the difference is not significant in the social-science group. This result indicates that the holistic scores can be affected by students' academic background, interest, and experiences to the given topic.

4. The two native-language groups' essay quality scores are significantly affected by the two different scoring methods. The difference between the two groups' scores are more significant when their essays are rated by the holistic scoring method.

The above results may be further clarified by the characteristics of the holistic scoring method itself. In the holistic scoring method, student essays are rated globally according to what the student has been able to perform compared with what the other students have been able to perform. The major function of the holistic scoring method is to separate the better performers from the poor performers by rank ordering the essays; hence, holistic scoring provides little information about the overall quality of an individual student's writing.

The results of this study raises one important issue, that is, the two scoring methods measure or emphasize different aspects of writing skills. The results of investigation on this issue is presented in the next section.

Correlations Among Different Measures of Writing Skill

In the preceding two sections of this chapter, the results of the investigation of topic effects were presented for syntactic characteristics and quality of students' essays written for two different writing tasks. In this section, in order to clarify and synthesize the findings of the preceding two sections, the results of the relationships between different measures of writing skill are presented in terms of simple correlations and multiple correlations in the following order: (a) the correlation between syntactic characteristic measures and essay quality measures (the analytic scores and holistic score), (b) correlations between analytic subcriteria scores and holistic score, and (c) correlations between writing skill measures (both quality measures and syntactic characteristic measures) and GRE Verbal scores. These three correlations were investigated separately for the two topics by using correlational analyses and multiple regression analyses methods. In addition, the relation was investigated through confirmatory factor analysis method.

The purpose of this section is twofold: (a) to examine the relationships between different measures of writing skill, and (b) to examine the topic effects on those relationships.

Correlations Between Syntactic Characteristics and Quality Scores

Correlational analysis. Table 11 gives the simple correlations of the six elaboration length variables and the syntactic complexity variables with the two quality variables for each of the two topics. The six elaboration length variables are the total number of words (TW), the total number of T-units (TU), the number of words in openers (WO),

the number of words in interrupters (WI), the number of words in closers (WC), and the number of words in free modifiers (WF; sum of WO WI WC). The five syntactic complexity variables are mean T-unit length (MT), mean opener length (MO), mean interrupter length (MI), mean closer length (MC), and mean free modifier length (MF; sum of MO MI MC). The two quality variables are the analytic total scores, and holistic scores. Table 11 shows that the elaboration length variables are more closely related to essay quality variables than are the syntactic complexity variables. This may indicate that the quality ratings of student essay are more significantly affected by the elaboration length variables than by the syntactic complexity variables. Among the five elaboration length variables, the total number of words (TW) is most highly correlated with essay quality variables, and the number of words in closers (WC) and the number of words in free modifiers (WF) are more highly correlated with essay quality than with the number of T-units.

Among the five syntactic complexity variables, the mean T-unit length (MT) is the highest. This suggests that the mean T-unit length is the most important, and mean closer length (MC) and mean free modifier length (MF) is the next important element in the quality rating. The correlation of mean closer length (MC) with essay quality is consistently higher across two topics than the mean opener length (MO) and mean interrrupter length (MI). The correlation between mean opener length (MO) and quality is not significant at alpha equal to .05 level. This may support several other researchers' findings that among mean opener length, mean interrupter length and mean closer length, mean closer length is the most important index of syntactic maturity (Christensen, 1968; Watson 1983 etc.).

Table 11

Correlations between Syntactic Characteristic Variables with Quality Scores

Variable	Analytic scores		Holistic scores	
	"Space"	"Farm"	"Space"	"Farm"
(Elaboration length)				
TW	.80	.79	.75	.61
TU	.45	.49	.36	.28
WO	.32	.20	.35	.20
WI	.38	.38	.38	.45
WC	.56	.47	.54	.47
WF	.64	.51	.64	.54
(Syntactic complexity)				
MT	.45	.37	.48	.42
MO	.10 (ns)	.13 (ns)	.16 (ns)	.04 (ns)
MI	.34	.31	.33	.40
MC	.37	.24	.38	.35
M F	.40	.18 (ns)	.43	.32

Note. The variable names are as follows:

TW: Total number of Words

TU: total number of T-Units

WO: number of Words in Openers

WI: number of Words in Interrupters

WC: number of Words in Closers

WF: number of Words in Free modifiers

MT: Mean T-unit length

MO: Mean Opener length

MI: Mean Interrupter length

MC: Mean Closer length

MF: Mean Free modifier length

In a comparison of the two types of topics, the correlations for the two topic data are similar. This fmding may indicate that the ratings of the "Space" topic essays are equally affected by the syntactic characteristics of the essays as the ratings of the "Farm" topic essays. In comparing the two types of scoring methods, the correlations of the elaboration length variables with analytic scores are higher than the correlations of the elaboration length variables with the holistic scores. This implies that the holistic scoring method would be less affected by the elaboration length of the essays than the analytic scoring method.

Multiple regression analysis. The above correlational analyses presented the degree of the relationship between the syntactic characteristic variables and the essay quality variables, under the assumption that the relationships are linear. However, that the relationships found in simple correlation are independent of other relationships is not clear from simple correlation because many of the variables used in the correlational analysis may be intercorrelated. The relationship may be better explained when they are analyzed simultaneously in terms of multiple regression analysis. In the stepwise multiple regression of this study, two quality variables (analytic scores and holistic scores) are dependent variables, and the other syntactic characteristic variables are independent variables. The stepwise multiple regression selects the single best predictor of quality scores first, and then the second best predictor in combination with the first , and so on until no more significant ($p < .05$) contribution is added to the prediction beyond the previous variables already entered.

Table 12 presents the results of the multiple regression analysis in predicting analytic total scores from the syntactic characteristic variables. The syntactic characteristic variables are entered separately as four groups: the elaboration length variables, the syn-

tactic complexity variables, the combined variables, and the combined variables without the total number of words variable.

Table 12 shows information about partial **R** square, **R** square, and standardized regression coefficient beta. The table includes only significant ($p < .05$) predictor variables. The syntactic complexity variables are least effective as a predictor of the analytic total scores. The elaboration length variables explain 65% of the total variance in the "Space" topic analytic scores and 63% of the total variance in the "Farm" topic analytic scores; the syntactic complexity variables explain 23% of the total variance for the "Space" topic and 25% of the total variance for the "Farm" topic.

Among the six elaboration length variables, the total number of words (TW) and total number of T-units (TU) variables make a significant contribution to the prediction of the analytic scores. The total number of words alone explained 64% of the total variance in the "Space" topic analytic scores, and 61% of the total variance in the "Farm" topic. These results suggest that the amount of semantic units and main idea units is highly correlated with the quality of the essay.

Among the five syntactic complexity variables, the mean T-unit length (MT) variable makes the most significant contribution to the prediction of the analytic scores in the both topics. The mean T-unit length explained 21% of the total variance for the "Space" topic, and 14% of the total variance for the "Farm" topic. The next most important predictor was mean closer length (MC) for the "Space" topic, and mean opener length (MO) for the "Farm" topic.

Table 12

Multiple Prediction of Analytic scores from Syntactic Characteristics

	"Space" topic				"Farm" topic			
Step	Variable	partial R^2	R^2	beta	Variable	partial R^2	R^2	beta
(Elaboration length variables only)								
1	TW	.635	.635	.833	TW	.608	.608	1.064
2	TU	.018	.653	-.136	TU	.025	.633	-.273
(Syntactic complexity variables only)								
1	MT	.205	.205	.453	MT	.138	.138	.542
2	MC(ns)	.022	.227	.167	MO	.111	.249	-.374
(Elaboration length and syntactic complexity variables combined)								
1	TW	.635	.635	.833	TW	.608	.608	.806
2	TU	.018	.653	-.136	MT	.031	.639	.202
3	MC(ns)	.005	.659	.098	WO	.028	.668	-.187
(Combined variables without TW variable)								
1	WF	.408	.408	.648	WF	.263	.263	.245
2	TU	.074	.482	.341	TU	.136	.399	.578
3	MT	.110	.592	.601	MT	.205	.605	.799
4	MF	.025	.617	-.550	MF	.024	.628	-.424

Note. The variable names are as follows:

TW: Total number of Words

TU: total number of T-Units

WO: number of Words in Openers

WI: number of Words in Interrupters

WC: number of Words in Closers

WF: number of Words in Free modifiers

MT: Mean T-unit length

MO: Mean Opener length

MI: Mean Interrupter length

MC: Mean Closer length

MF: Mean Free modifier length

When the complexity variables are added to the elaboration length variables in the prediction of the analytic scores, the total number of words variable is still the most important predictor in both topics, and the next two important predictors are the number of T-units (TU) and the number of words in closers (WC) for the "Space" topic, and mean T-unit length (MT) and the number of words in openers (WO) for the "Farm" topic. The three variables in each topic explained 66% of total variance in the "Space" topic analytic scores and 67% of total variance in the "Farm" topic analytic scores. In both topics, the free modifier length variables do not make a significant contribution to the prediction. Of particular interest is that the combined variables explain almost the same amount of variance as do the elaboration length variables only.

The above results indicate that the essay length variables correlate more highly with the essay quality variables than do the syntactic complexity variables. These results are consistent with other researchers' findings (Breland, 1983; Hendrickson, 1980). The reason may be explained by the fact that the syntactic complexity variables measure the syntax level skill, whereas the elaboration length variables reflect the content and structure of the overall essay.

When the total number of words variable is deleted from the combined variables, the number of words in free modifiers (FW) variable is the most important predictor for the analytic scores in the both topics. next three significant contributors are the number of T-units (TU) , mean T-unit length (MT), free modifier length (MF) for both topics. These four variables explain 62% of the total variance in the "Space" topic analytic scores and 63% of the total variance in the "Farm" topic analytic scores. Thus, these four variables for each topic explain almost the same amount of variance in the analytic scores as does the total number of words variable alone. This fmding indicates that the

total number of words variable is obviously correlated with various other attributes of writing skill.

In comparing the two types of topics, the elaboration length variables, the syntactic complexity variables, and the combined variables in the the "Space" topic make almost the same degree of contribution to the prediction of the analytic scores as do those variables in the "Farm" topic. This finding indicates that the relations between syntactic characteristics and analytic scores are consistent across the two topics.

Table 13 presents the results of the multiple prediction of the holistic scores from syntactic characteristics. As in the prediction for analytic scores, the table shows that the syntactic complexity variables are less effective than the elaboration length variables in the prediction of the holistic scores. While the elaboration length variables explain 60% of the total variance in the "Space" topic and 46% of the total variance in the "Farm" topic, the syntactic complexity variables explain only 24% of the total variance in the "Space" topic and 25% of the total variance in the "Farm" topic.

Among the elaboration length variables, the total number of words (TW) variable is the most important predictor for the holistic scores. This variable explained 56% of the total variance in the "Space" topic holistic scores and 37% of the total variance in the "Farm" topic holistic scores. The number of T-unit variable is the next important predictor of the holistic scores in both topics.

Among the five syntactic complexity variables, the mean T-unit length (MT) variable made the most significant contribution to the prediction of the holistic scores for both topics. As in the case of prediction for the analytic scores, this may indicate that among

Table 13

Multiple Prediction of Holistic Scores from Syntactic Characteristics

	"Space" topic				"Farm" topic			
Step	Variable	partial R^2	R^2	beta	Variable	partial R^2	R^2	beta
(Elaboration length variables only)								
1	TW	.559	.559	.936	TW	.368	.368	.772
2	TU	.041	.600	-.276	TU	.075	.443	-.301
(Syntactic complexity variables only)								-
1	MT	.226	.226	.475	MT	.181	.181	.562
2	MC(ns)	.017	.243	.147	MO	.072	.253	-.302
(Elaboration length and syntactic complexity variables combined)								
1	TW	.559	.559	.833	TW	.368	.368	.494
2	TU	.041	.600	-.136	MT	.077	.446	.372
3	MT(ns)	.004	.604	.098	OT	.020	.465	-.162
(Combined variables without TW variable)								
1	FW	.412	.412	.607	FW	.291	.291	.155
2	FT	.042	.454	.521	FT	.026	.317	.510
3	MT	.104	.557	-.880	TU	.106	.423	.437

Note. The variable names are as follows: TW: Total number of Words

TU: total number of T-Units

WO: number of Words in Openers

WI: number of Words in Interrupters

WC: number of Words in Closers

WF: number of Words in Free modifiers

MT: Mean T-unit length

MO: Mean Opener length

MI: Mean Interrupter length

MC: Mean Closer length

MF: Mean Free modifier length

the six syntactic complexity variables, the mean T-unit length is the most important predictor for a quality ratings of student essays. The mean T-unit length explains 23% of the total variance for the "Space" topic, and 18% of the total variance for the "Farm" topic. The next most important predictor is the mean closer length (MC) for the "Space" topic, and the mean opener length (MO) for the "Farm" topic. However the mean closer length variable does not make a statistically significant ($p < .05$) contribution to the prediction.

When the combined variables (elaboration length variables and complexity variables) are entered into the multiple regression, the total number of words variable is the most important predictor for both topics, and the next two important predictors are the number of T-units (TU) and the mean T-unit length (MT) for the "Space" topic, and the mean T-unit length (MT) and the mean opener length (OT) for the "Farm" topic. The three variables in each topic explain 60% of the total variance in the "Space" topic holistic scores and 47% of the total variance in the "Farm" topic holistic scores. For the holistic scores, the combined variables explain almost the same amount of variance as does the elaboration length variables only. These results are consistent with the multiple prediction of the analytic scores. The results support several researchers' findings that the holistic scores correlates more highly with essay length than with syntactic complexity variables (Grobe, 1981; Stewart & Grobe, 1979). They reported that the holistic ratings focussed more on the amount of ideas generated in the essay than on the syntactic level skill or mature style of syntax generated in the essay.

When the total number of words variable is deleted from the combined variables, the number of words in the free modifiers (FW) variable is the most important predictor for the holistic scores for both topics. The next two significant contributors are mean free

modifier length (FT) and mean T-unit length (MT) for the "Space" topic, mean free modifier length and number of T-units (TU) for the "Farm" topic. These three variables for each topic explain 56% of the total variance in the "Space" topic holistic scores and 42% of the total variance in the "Farm" topic holistic scores. Once again, the results indicate that the total number of words variable is obviously correlated with several other attributes of writing skill.

In comparing the two types of topics, the contribution of the syntactic characteristic variables is higher for the the "Space" topic holistic scores than for the "Farm" topic holistic scores when the holistic scores are predicted from elaboration length variables and combined variables of elaboration length with complexity. However, when the holistic scores are predicted from the syntactic complexity variables, the two topics do not show any significant differences. This finding may imply that the prediction of the quality scores from syntactic complexity is not influenced significantly by different types of topic.

In comparing the two types of quality scores, the syntactic characteristic variables made a greater contribution to the analytic scores than to the holistic scores. This indicates that the analytic rating may be influenced more by syntactic characteristics, (especially the length of elaboration) than is the holistic rating. This result may provide important information for the two scoring methods. According to the result, the analytic total scores reflect more highly the amount of ideas generated in the essay than the holistic scores.

Confirmatory factor analysis. To examine the degree of relationship between the quality measures and the syntactic complexity measures, an analysis was conducted

based on the confirmatory factor analysis model using the "Space" topic data. This study analyzed a model in which content/thinking, organization, style/tone, and the holistic score variables were assumed to be indicators of an essay quality factor, and mean T- unit length and mean closer length variables were assumed to be indicators of a syntactic complexity factor. The primary interest of the analysis was in estimating the true correlation between the essay quality factor and the syntactic complexity factor.

The model was tested and estimated using a computer program LISREL (version 6.6) for the maximum likelihood estimates. The results are presented in Table 14. According to the results, the measure of fit indices indicates a good fit of the model (chi-square = 11 df= 8 p = .847; root mean square residual= .019; adjusted goodness of fit index = .965). T-values for all parameters also indicate that all parameters are highly significant (all T-values are larger than five).

The factor intercorrelation (estimated correlation between two latent variables) is .497 with a standard error of .09. Even though the intercorrelation is higher than the observed correlations between essay quality variables and syntactic complexity variables, it is still only moderate. This modest interfactor correlation indicates that the syntactic complexity variables measure different attributes of writing skill than the essay quality variables because the two factors do not share a large amount of common variance. Of the indicators of the syntactic complexity factor, the mean closer length variable has a lower factor loading than the mean T-unit length variable. This result is consistent with the finding of the multiple regression analysis, suggesting that the mean T-unit length is the most important predictor of essay quality among syntactic complexity variables.

Table 14

Relation between Essay Quality and Syntactic Complexity: Maximum Likelihood Estimates

Factor Loading

Variables/ Factor	Essay quality	Syntactic complexity
Holistic score	.868	0
Content/thinking	.969	0
Organization	.971	0
Style/tone	.943	0
Mean T-unit length	0	.966
Mean closer length	0	.736

Factor Intercorrelation

	Essay quality	Syntactic complexity
Essay quality	1.000	
Syntactic complexity	.497	1.000

Summary and discussion. The preceding results of the investigation on the relationships between syntactic characteristics and quality ratings show the following:

1. Of the two types of topics, the amount of prediction from syntactic complexity variables is similar for both the "Space" topic end for the "Farm" topic.

2. Of the two types of scoring methods, the amount of prediction from the elaboration length variables is higher for the the analytic scores than for the the holistic scores.

3. Of the elaboration length variables and the syntactic complexity variables, the elaboration length variables make a more significant contribution to the prediction of both the analytic scores and the holistic scores than do the syntactic characteristic variables.

4. Among the elaboration length variables, the total number of words variables is the most important predictor for both the analytic and holistic scores.

5. Among the syntactic complexity variables, the mean T-unit length variable is the most significant contributor for the prediction of both holistic and the analytic scores. However, the amount of prediction is smaller than that of the total number of words variable (the R square range is from .14 to .23).

6. Among all the syntactic characteristic variables, the total number of words variable is still the most important predictor, and the next important predictors are number of T-units or mean T-unit length, depending on the different topic type.

7. When the total number of words variable is deleted from the combined variable in the multiple regression, the elaboration length variables are still a more significant predictor than the syntactic complexity variables. In this case, the total amount of prediction made from all significant predictors is almost the same as the amount of prediction made from the single total number of words variable.

The above findings may indicate that the relationships between syntactic characteristics and the quality ratings could be affected by different types of topics and scoring methods. Although the topic effect and scoring method effect on the relationships were not directly investigated, several researchers suggest that the relationships can be affected by the different discourse modes, the audience, or grade level (e.g., Stewart and Grobe 1979; Breland 1983 etc.). The results of this study also support several other researchers' findings that the essay length variable is the single most important factor for the quality of the essays among several kinds of syntactic characteristic variables. The results of a confirmatory factor analysis indicate that the interfactor correlation between essay quality and syntactic complexity is only moderate, suggesting that the two types of variables measure different attributes of writing ability.

Correlations Between Analytic and Holistic Ratings

Correlational analysis. Table 15 gives the simple correlations of the thirteen analytic subcriteria and three main criteria scores with the holistic scores for each of the two topics. The thirteen subcriteria are adequacy of information, richness of information, relationships, inferences, syntheses, evaluation, alternatives, framing, grouping, unity, objectivity, tentativeness, and metalanguage. The three main criteria are content/thinking, organization, and style/tone.

Table 15 presents the significant correlations between the holistic scores and the analytic subscale scores for the two topics. The significant correlations indicate that each of the analytic subcriteria and main criteria commonly have certain attributes of writing skill as does the holistic scale. The correlations between analytic subcriteria and the holistic scores for the "Space" topic data are higher than the correlations for the "Farm" topic. This finding may imply that the relationship between holisting ratings could be changed by different types of writing tasks. In addition, the holistic ratings might emphasize different attributes of writing skills according to different types of writing tasks.

Among the thirteen analytic subcriteria, the richness of information, relationships, and unity subcriteria scores were highly correlated with the the holistic scores for the "Space" topic data, whereas for the "Farm" topic data, the framing, unity, and richness of information subcriteria scores are correlated most highly with the holistic scores.

Of the three analytic main criteria scores, the content/thinking main criteria scores correlate highest with the holistic scores for the "Space" topic data. The organization main criterion scores correlate highest with the holistic scores for the "Farm" data. In

Table 15

Correlations Between Analytic Scores and Holistic Scores

a. correlations of analytic scores with holistic scores

Analytic subscale	"Space" holistic score	"Farm" holistic score
(Analytic subcriteria)		
Adequacy of information	.70	.69
Richness of information	.79	.73
Relationships	.80	.68
Inferences	.75	.68
Synthesis	.68	.66
Evaluation	.74	.65
Alternatives	.70	.58
Framing	.73	.75
Grouping	.67	.72
Unity	.81	.73
Objectivity	.72	.67
Tentativeness	.71	.71
Metalanguage	.69	.67
(Analytic main criteria)		
Content/thinking	.82	.74
Organization	.80	.77
Style/tone	.76	.78
Analytic total scores	.82	.79

b. correlations between analytic total and holistic scores

variables	1	2	3	4
1 Analytic total ("Space")	1.00			
2 Analytic total ("Farm")	.80	1.00		
3 Holistic scores ("Space")	.82	.74	1.00	
4 Holistic scores ("Farm")	.83	.79	.86	1.00

the total the analytic scores, the correlation is .82 for the the "Space" data, and .79 for the "Farm" data.

Multiple regression analysis. The above correlational analysis presented the degree of the relationships between the analytic scores and the holistic scores for each of two topics. To examine the relationships simultaneously in terms of stepwise multiple regression, two groups of variables are entered separately for the two topics into stepwise multiple regression equation: (a) thirteen analytic subcriteria scores, and (b) three analytic main criteria scores. In the equation, the dependent variable is the the holistic scores of each topic.

Table 16 presents the results of the multiple regression analysis in predicting the holistic scores from analytic subcriteria and main criteria scores. The table shows information about the partial R square,**R** square,andstandardizedregressioncoefficientbeta. The table includes only significant ($p < .05$) predictor variables.

For the "Space" topic data, three analytic subcriteria scores--unity, richness of information, and alternatives--make significant contributions to the prediction of the holistic scores. These three variables explain 72 % of the total variance in the "Space" topic holostic scores. The unity subcriteria scores alone accounts for 65 % of the total variance in the the holistic scores. Among three analytic main criteria scores of the "Space" topic, content/thinking and organization main criteria scores contribute significantly to the prediction of the holistic scores. These two variables explain 68 % of the total variance in the holistic scores. The content/thinking main criteria alone explains 67% of the total variance in the holistic scores.

Table 16

Multiple Prediction of Holistic Scores from Analytic Scores

"Space" topic

Step	Variable	partial R^2	R2	beta
(Analytic subcriteria)				
1	Unity	.650	.650	.421
2	Richness of information	.060	.710	.344
3	Alternatives.	.012	.722	.160
(Analytic main criteria)				
1	Content/thinking	.665	.665	.523
2	Organization	.013	.679	.315

"Farm" topic

Step	Variable	partial R^2	R2	beta
(Analytic subcriteria)				
1	Framing.	.569	.569	.374
2	Metalanguage.	.060	.628	.325
3	Objectivity.	.018	.646	.211
(Analytic main criteria)				
1	Style/tone	.615	.615	.486
2	Organization	.025	.630	.336

For the "Farm" topic, of the thirteen analytic subcriteria, the framing, metalanguage, and objectivity variables are statistically significant predictors for the "Farm" topic holistic scores. These three variables explain 65% of the total variance in the "Farm" topic holistic scores. Among the three main criteria variables, style/tone and organization are significant contributors to the prediction, explaining 64% of the total variance in the holistic scores.

The above results indicate that the relationship between the analytic ratings and the holistic ratings varies from topic to topic. While the holistic ratings for the "Space" topic rely heavily on the dimension of the content/thinking (especially the richness of information and alternatives criteria) and organization (especially the unity criterion), the holistic ratings for the "Farm" topic focus on the dimension of the style/tone (especially the metalanguage and objectivity subcriteria) and organization criteria (especially the framing subcriterion). This result may be related to the finding on the scoring method and topic interaction effect in the second section of this chapter. The higher scores on the "Farm" topic may be due to the criterion used for holistic judgment for the essay, although the criterion was not explicitly established. For example, in the analysis of topic effects on the analytic scores, the objectivity subcriterion scores of the "Space" topic were significantly lower than the scores of the "Farm" topic essays.

Confirmatory factor analysis: The multiple regression analyses show the relationships between the analytic scores and the holistic scores. However, the analyses do not show whether the two groups of analytic main criteria scores for each topic measure different traits of writing skills. If the two groups of analytic main criteria scores are unique for each topic, there appears to be two distinctive factors, and the two factor model may explain the relationship between two topics. For this purpose, a

confirmatory factor analysis was conducted. The three analytic main criteria variables for the "space" topic were assumed to be indicators of writing skill for a general topic (a topic with general information provided) factor, and another three main criteria variables for the "Farm" topic were assumed to be indicators of writing skill for a specific topic (a topic with narrowly defined specific information provided) factor.

Table 17 presents the maximum likelihood estimates of the model. The results indicate a good fit of this model yielding nonsignificant chi-square and other indices of good fit (chi-square= 12.53 df= 8 p=.129; root mean square residual= .02; adjusted goodness of fit index= .89).

All factor coeficients are very high (from .928 to .972), and among those coeficients, the organization criterion has the largest factor loading for both topics. Within each factor, the three analytic main criteria measure equally and effectively the latent traits of writing skill; hence, the criteria may not be disdinguished from each other.

The interfactor correlation is .814 with a standard error of .038. The factor inter- correlation is almost invariant when the sample size is increased from 96 to 212 (the increased sample yielded a factor correlation of .799 with a standard error of .027). The correlation corrected for attenuation between factors tends to be high because the maximum likelihood factor analysis model consider all unique variance under the constrained model to be error variance. Neverthless, the factor correlation of .814 is relatively high when comparing with the essay reliability estimation of previous studies (Breland, 1983; Coffman 1971b; Godshalk et al., 1966; Werts et al., 1980). Therefore, the interfactor correlation may show that students' writing performance on the two topics of this study does not seriously fluctuate from one topic to the other topic. This may further indicate

Table 17

Relationship between General Topic and Specific Topic: Maximum Likelihood Estimates

Factor Loading

Variables/ Factor	General topic	Specific topic
Content/thinking (Space)	.966	0
Organization (Space)	.972	0
Style/tone (Space)	.946	**0**
Content/thinking (Farm)	0	.955
Organization (Farm)	0	.960
Style/ tone (Farm)	0	.928

Factor Intercorrelation

	General topic	Specific topic
General topic	1.000	
Specific topic	.814	1.000

that, in a testing situation, the two topics (general vs. specific) may provide redundant information about student writing ability.

In addition to the relationship between the topics, the relationship between the thirteen analytic subcriteria and the analytic main criteria was investigated using the original sample's total analytic scores of the two topics. First, a principal factor analysis was conducted on the basis of an oblique rotation method (Promax) using SAS computer program (Cureton & Mulaik, 1975; Harmon, 1976; Kaiser & Cerny, 1979). The result indicates that the data of this study is appropriate for the common factor model (Kaiser's measure of sampling adequacy= .97), and the best number of common factors for the thirteen subcriteria is three. There are three large positive eigenvalues that to-

gether account for more than 100 % of the common variance. The rotated factor pattern of the thirteen subcriteria shows that the three factors are approximated to the three analytic main criteria (see Appendix B for the rotated factor pattern). However, some of the subcriteria relate more closely to the other main criterion than the original main criterion of the analytic scoring scheme. For example, the metalanguage subcriterion is more closely related to the organization criterion (factor loading= .552) than to the style/tone main criterion (factor loading = .252). Based on this result, a confirmatory factor analysis was conducted to examine the true correlation among the three main criteria. The result indicates a reasonable fit of the model (see Appendix B for the result). An interesting finding is that the interfactor correlations are higher than .97. These high correlations show that the three factors are not distinguishable from each other; hence, the three factors are actually the same index of writing skill. This result may be due to the following two facts: (a) the characteristics such as content, organization, and style are inextricable in a single piece of writing, and (b) in the analytic scoring method, there is a danger of overlooking particular elements of writing skill because of a previously formed general impression (so called the halo effect).

Summary and discussion. These results clearly show two major points. First, the patterns of correlation between the holistic scores and the analytic scores are significantly affected by the different types of topics. The holistic scores of the "Space" topic are highly correlated with the analytic content/thinking scores, whereas the "Farm" topic holistic scores are highly correlated with the analytic style/tone scores. The analytic organization scores are highly correlated with the holistic scores of both topics. The results imply that the holistic ratings of the "Space"topic emphasize the content/thinking aspects of writing skills (especially abundant information and alternative views' of thinking), and the holistic ratings of the "Farm" topic emphasize the style/tone aspect

of writing skills (especially the objective position on the given theme and the use of appropriate markers indicating the relationships of ideas).

Second, based on the assumption that the general topic and the specific topic measure different aspects of writing skill, the model shows an interfactor correlation of .814 with standard error of .038. The correlation is relatively high when compared with the essay reliability estimation of previous studies, indicating that the analytic scores of the two topics (general vs. specific) do not seriously change from one topic to the other as other researchers suggest (Odell et al., 1978; Breland, 1983; Meredith & Williams, 1984).

Correlations Between Writing and Verbal Skill Measures

Correlational analysis. Table 18 gives simple correlations between two different types of writing skill measures (quality ratings and syntactic characteristics) and GRE verbal section scores. The GRE verbal scores was based on the sum of sentence completion, discrete verbal, and reading comprehension test scores. The correlations are presented for each of the two topics.

According to the results, among essay quality measures, the holistic scores correlates most strongly with the verbal scores for both topics. Of the analytic main criteria scores, the style/tone scores correlate best with the verbal scores for the both topics. The correlations of essay quality variables with verbal scores show a similar pattern for both topics. This indicates that the relationship between essay quality measures and GRE verbal scores is consistent with the two different types of topics.

Table 18

Correlations between GRE Verbal Score and Measures of Writing Skill

Writing skill measures	GRE verbal scores (with Space" data)	GRE verbal scores (with "Farm" data)
Holistic scores	.82	.84
(Analytic scores)		
Content/thinking	.75	.68
Organization	.75	.71
Style/tone	.76	.74
Analytic total scores	.78	.73
(Syntactic characteristics)		
TW	.73	.57
TU	.35	.22
WC	.51	.48
WF	.60	.52
MT	.48	.42
MC	.35	.37
MF	.41	.31

Note. The variable names are as follows:

TW: Total number of Words

TU: total number of T-Units

WC: number of Words in Closers

WF: number of Words in Free modifiers

MT: Mean T-unit length

MC: Mean Closer length

MF: Mean Free modifier length

Of the syntactic characteristic variables, the total number of words variable correlates highest with the verbal scores, and the correlation of number of words in free modifiers with the verbal scores is next most high. This indicates that the elaboration length variables are more highly correlated with the verbal scores than are the syntactic complexity variables.

Among syntactic complexity variables, the mean T-unit length variable correlates highest with the GRE verbal score. The correlations of syntactic characteristic variables with the verbal scores are higher for the "Space" topic than for the "Farm" topic. The results are consistent with the finding regarding the relationships between syntactic characteristic variables and essay quality variables.

Multiple regression analyses. Table 19 presents the results of the multiple regression analysis in predicting GRE verbal scores from writing skill measures.

In the stepwise multiple regression, four groups of variables were entered separately: (a) the analytic main criteria variables only, (b) the adding of the holistic scores variable to the analytic main criteria variables, (c) and syntactic characteristic variables only, (d) the adding the analytic main criteria variables to the syntactic characteristic variables. Among the analytic main criteria variables, the style/tone is the most important predictor for the GRE verbal scores for both topics. This may indicate that the style/tone criteria and the GRE verbal section measures similar attributes.

When the the holistic score variable is added to the main criteria variables, the holistic score and the style/tone score variables are significant contributors for the prediction. These two variables explain over 70 % of the total variance in the GRE

Table 19

Multiple Prediction of GRE Verbal Score from Writing Skill Measures

"Space" topic

Step	Variable	partial R^2	R2		beta
(Analytic main criteria)					
1	Style/tone	.578	.578		.443
2	Organization	.022	.600		.349
(Adding the holistic scores variable to the main criteria)					
1	Holistic scores	.676		.676	.578
2	Style/tone	.043	.719		.320
(Syntactic characteristic variables)					
1	TW	.540	.540		.924
2	TU	.042	.582		-.278
(Adding syntactic characteristic variables to the analytic scores)					
1	style/tone	.578	.578		.425
2	TW	.047	.625		.532
3	TU	.042	.645		-.198

"Farm" topic

Step	Variable	partial R^2	R2	beta
(Analytic main criteria)				
1	Style/tone	.615	.615	.486
2	Organization	.025	.639	.336
(Adding the holistic scores variable to the main criteria)				
1	Holistic scores	.713	.713	.696
2	Style/tone	.014	.727	.189
(Syntactic characteristic variables)				
1	TW	.325	.325	.942
2	TU	.104	.429	-.491
(Adding syntactic characteristic variables to the analytic scores)				
1	Style/tone	.540	.540	.631
2	TW	.367	.577	.217

Note. The variable names are as follows:

TW: Total number of Words

TU: total number T-Units

verbal scores. the holistic score is the most powerful predictor for the verbal scores among all variables entered in the multiple regression. The variable alone explain 68 % of the total variance in the verbal scores for the "Space" topic, and 71 % of the total variance for the "Farm" topic.

Of the eleven syntactic characteristic variables entered, only two elaboration length variables--total number of words and number of T-units--make a significant contribution to both topics; however, the amount of contribution is higher for the "Space" topic than for the "Farm" topic. When the syntactic characteristic variables are added to the analytic main criteria, the style/tone main criterion and total number of words variables make a significant contribution in the prediction of the GRE verbal scores for both topics.

In a comparison of the two topics, the patterns of prediction are similar for both topics. The most important result is that, of the essay quality and syntactic characteristic variables, the holistic scores and style/tone scores variables share a larger common variance with the GRE verbal scores than do the other variables.

Confirmatory factor analysis. To further investigate the relationships between the GRE verbal scores and the essay quality scores, a confirmatory factor analysis was conducted. In the analysis, three types of GRE verbal scores (sentence completion, discrete verbal, and reading comprehension) were assumed to be indicators of general verbal skill, and four types of essay quality measures were assumed to be indicators of writing skill.

Table 20

Relation between Writing Skill and General Verbal Skill: Maximum Likelihood Estimates

Factor Loading

Variables/ Factor	Writing skill	Verbal skill
Analytic total (Space)	.891	0
Holistic scores (Space)	.914	0
Analytic total (Farm)	.842	0
Holistic scores (Farm)	.937	0
Sentence completion	0	.910
Discrete verbal	0	.904
Reading comprehension	0	.912

Factor Intercorrelation

	Writing skill	Verbal skill
Writing skill	1.000	
Verbal skill	.923	1.000

Table 20 presents the maximum likelihood estimates of factor coefficients and factor intercorrelation.

The results indicate a reasonable fit of the model, yielding root mean square residual of .025 and adjusted goodness of fit index of .829 although the chi-square is significant (chi-square= 33.23 df= 13 p= .002). The factor intercorrelation is .923 with a standard error of .022. The factor correlation .923 represents the correlation between the writing skill factor and the general verbal skill factor when both sets of measures are corrected for unreliability. The correlation for attenuation between factors in the maximum likelihood estimate tends to be high because the maximum likelihood estimation method considers all unique variance under the constrained model to be error variance

(Joreskog & Sorbom 1979; Rock, Werts, & Grandy, 1983). Even considering this fact, the interfactor correlation .923 with a standard error of .623 is comparatively high. This indicates that the two factors share a large common variance.

However, in the above model, there is a pattern of normalized residuals associated with the reading comprehension variable that is larger than two in magnitude. In addition, the largest modification index was found in the variable (modification index for reading comprehension= 10.87). Considering this fact and the nonsignificant chi-square of the model (chi-square= 32.2 df= 13 p= .002), the initial model was revised by setting the factor loading of the reading comprehension variable free to find a better model fit.

Table 21 presents the maximum likelihood estimates of the revised model. The revised model reduced the chi-square from 33.2 to 22.3, and also reduced the index of root mean square residual from .022 to .017, and the adjusted goodness of fit index was .870. The interfactor correlation was also reduced from .923 to .885. This may indicate that the shared common variance of the two factors can be explained by the reading comprehension variable to a certain degree.

The factor loading of the reading comprehension variable was .471 on the writing skill factor, and .461 on the verbal skill factor. This results suggest that the variance of the reading comprehension variable is equally and highly correlated with both the writing skill factor and the verbal skill factor. The above analysis is based on a small sample size (n= 96); hence, the results should be interpreted with limitations.

The above analysis presents only the relationship between essay quality measures and verbal skill measures. However, several writing researchers emphasize the relationship

Table 21

Relationship between Writing Skill and General Verbal Skill: Maximum Likelihood Estimates--Revised Model.

Factor Loading

Variables/ Factor	Writing skill	Verbal skill
Analytic total (Space)	.890	0
Holistic scores (Space)	.913	0
Analytic total (Farm)	.843	0
Holistic scores (Farm)	.938	0
Sentence completion	0	.938
Discrete verbal	0	.919
Reading comprehension	.471	.461

Factor Intercorrelation

	Writing skill	Verbal skill
Writing skill	1.000	
Verbal skill	.885	1.000

between writing skill and reasoning skill (Applebee, 1984; Emig, 1971; Flower and Hayes, 1982, 1984). It has been widely accepted that writing skill requires reasoning skill, problem solving strategies, and higher order thinking skills. Therefore, further analysis is needed to examine the relationships between writing, verbal, and reasoning skills. In this analysis, the GRE logical reasoning scores and the analytical reasoning scores were assumed to be indicators of reasoning skill, and the original ETS sample (n= 205) was used for this analysis because the small sample data (n= 96) of this study did not provide an acceptable fit.

The results are presented at Table 22, showing that the three factor model yields a reasonable fit (chi-square = 39.90 df= 24 p= .022; adjusted goodness of index= .927; root

Table 22

Relationship between Writing, Verbal, and Reasoning skill: Maximum Likelihood Estimates

Factor Loading

Variables/ Factor	Writing skill	Verbal skill	Reasoning skill
Analytic total (Space)	.888	0	0
Holistic scores (Space)	.819	0	0
Analytic total (Farm)	.841	0	0
Holistic scores (Farm)	.928	0	0
Sentence completion	0	.899	0
Discrete verbal	0	.863	0
Reading comprehension	0	.907	0
Analytical reasoning	0	0	.502
Logical reasoning	0	0	.944

Factor Intercorrelation

	Writing skill	Verbal skill	Reasoning skill
Writing skill	1.000		
Verbal skill	.898	1.000	
Reasoning skill	.773	.878	1.000

mean square residual= .017). Inspection of the factor intercorrelation shows that the writing skill factor has a higher relationship with the general verbal skill factor than with the reasoning skill factor. It also shows that the reasoning skill factor haS a higher relationship with the verbal skill factor than with the writing skill factor. This may indicate that writing quality measures depend more heavily on the general verbal skill factor than on the reasoning skill factor as found in the student essays although the low correlation between the two factors can be explained by the comparatively low factor loading (.502) in the analytical reasoning variable. However, the correlation between the writing skill

factor and the reasoning skill factor is .773 with a standard error of .046, indicating that the two factors have a large shared variance.

Summary and discussion. The above results suggest the following:

1. The correlations between the essay quality scores and the GRE verbal scores are stable across the two topics.

2. Of all the essay quality variables and syntactic characteristic variables, the holistic score is the most significant predictor of the GRE verbal scores. The result indicates that the holistic scores and GRE verbal scores share a large portion of variance.

3. Of the analytic scores variables, the style/tone main criterion scores make the most important contribution to the prediction of the verbal scores in both topics, suggesting that the ratings of style/tone criterion emphasize the same skill that the objective GRE verbal test measures.

4. Of the the analytic scores and syntactic characteristic variables, the style/tone scores variable and the total number of words variables make important contributions to the prediction of the verbal scores.

5. The interfactor correlation between the writing skill and the general verbal skill is very high, indicating that the two factors measure almost the same skills. The reading comprehension variable has almost the same degree of factor loading for both factors. The factor correlation between writing and verbal skill is higher than the factor correlation between writing and reasoning skill.

CHAPTER V
CONCLUSIONS AND IMPLICATIONS

This chapter presents a summary of procedures, the principal results, discussion and conclusions, and the implications of this study for theory, research, and assessment.

Summary of Procedures

In effort to determine whether two different topics (a writing task with general information provided and a writing task with narrowly defined specific information provided) will elicit different qualities and syntactic characteristics of writing from a sample of upper college level students, this study raised the following three major questions: (a) do the two different topics affect significantly students' writing performance in the syntactic characteristics and in the quality ratings in a testing situation? (b) are the topic effects consistent across different groups of writers (different native-language groups and different academic-major groups)? and (c) do the relationships among different measures of writing skill remain stable across the two different topics?

In conjunction with the above questions, a review of the related research was conducted for the major studies on assignment problems in writing assessment and for the major studies on scoring methods for writing assessment. The studies of writing- assignment problems were reviewed concerning both theoretical perspectives and empirical studies. Similarly, studies of scoring methods were reviewed, focussing on the analytic and the holistic methods with a consideration of the reliability problems. Through this review, the following three major points were identified: (a) the types of information provided in the writing assignment may play a significant role in the process

and production of the writing, (b) however, only limited empirical research exists to provide the evidence for the effect of the information type in the writing assignment, and (c) the different scoring methods may lead to different interpretations for the assignment effect on the writing performance. The results of this review provided a background for this study.

For the purposes of this study and as a result of the problems found in the review, ninety six upper level college students' writing samples were selected from the original ETS writing samples: (a) twenty-four students, native speakers of English, with a major in a hard science; (b) twenty-four students, native speakers of English, with a major in a social science; (c) twenty-four students, native speakers of Chinese, with a major in a hard science; (d) twenty-four students, native speakers of Chinese, with a social science major. Since each student wrote two essays on two different topics, the total number of writing samples for this study was 192. Four different kinds of writing skill measures were used: elaboration length, syntactic complexity, analytic scores, and holistic scores. For the elaboration length measure, the total number of words, the total number of T- units, and the total number of words in free modifiers were counted. For syntactic complexity, the mean T-unit length, the ratio of free modifiers, and the ratio of closers were counted. The analytic scores were generated for the original ETS writing samples by four raters using Purves's (1985) analytic scoring method. The analytic scoring scale includes three main criteria and thirteen subcriteria of writing skill. The holistic scores was generated by two raters for the original writing samples.

The characteristics of the two topics ("Space" and "Farm") were examined in relation to the elements of the writing assignment. The two topics were similar in the discourse mode, purpose, audience, cognitive-processing-demand level, and task-complexity level;

but the two topics were distinctive in the types of information (general vs. specific) provided. In the "Space" topic, students were required to compare and contrast the advantages and disadvantages of space exploration and to take a position; whereas, in the "Farm" topic, students were required to interpret the relationships among the three graphs showing the changes in farming patterns over a period of forty years in United States.

To investigate the topic effect, the topic and writer group interaction effect, and the relationships among the different measures of writing skill, two kinds of analysis methods were used: the repeated measure of analysis of variance method (based on the split plot factorial design) and the correlational analysis method. For the analysis of variance, four kinds of analytic scores, one kind of holistic score, three kinds of elaboration-length measures, and three kinds of syntactic-complexity measures were used as dependent variables. And the two topic types, the two native-language groups, and the two academic-major groups were used as dependent variables. When the topic effects on the scoring methods and on the analytic main criteria were investigated, the two scoring methods and the three analytic main criteria were used as dependent variables. All these dependent variables were assumed to be fixed factors in the SPF22.2 design (two between subject treatments and one within subject treatment), the SPF2.22 design, and the SPF2.23 design (one between subject treatment and two within subject treatments).

For the correlational analysis, the simple correlation, the stepwise multiple regression, and the confirmatory factor analysis were used to investigate the following three relationships: (a) correlations between the syntactic characteristic measures and the essay quality measures, (b) correlations between the analytic scores and the holistic scores, and (c) correlations between the writing skill measures and the general verbal skill measures.

Principal Results

In this section, the principal results of the investigation for the topic effects are presented in conjunction with the three major questions of this study: (a) the topic effects on writing performance, (b) the topic and writer group interaction effect on writing performance, (c) and the relationships between different measures of writing skill. In addition, the results of the investigation on the group differences and the reading consistency problems are presented.

Topic Effects

Significant topic effects were found in the elaboration-length measures, the analytic content/thinking scores, and the holistic scores. However, in the syntactic complexity measures and all the analytic scores except the content/thinking scores, the topic effects were not significant. The following is a brief summary of the findings on the topic effects (in order of significance):

1. The students generated significantly longer elaboration for the writing task with general information provided (the "general" writing task) than for the writing task with a narrowly defined specific information provided (the "specific" writing task).
2. The essays for the general writing task received significantly higher scores in the analytic content/thinking main criteria than the essays for the specific writing task. Of the seven subcriteria of the content/thinking dimension, significant topic effects were found in the richness of information, the evaluation, and the alternatives subcriteria.

3. The essays for the general writing task received significantly lower scores in the holistic rating than the essays for the specific writing task.

4. In the syntactic-complexity measures, the essays for the general writing task were similar to the essays for the specific writing task.

5. In the analytic organization, style/tone, and total analytic scores, both essays for the general writing task and for the specific writing task received almost the same degree of ratings.

Interaction Effects

The following is a summary of the major findings on the interaction effect of the native-language groups (NL) and topics, on the interaction effect of the academic-major groups (MG) and topics, on the interaction effect of the analytic main criteria and topics, and on the interaction effect of the scoring methods and topics.

1. The NLxtopic interaction effects were significant only in the two elaboration-length measures: the total number of words and the number of words in free modifiers. This result indicates that the native-English language group used a significantly larger number of words for the general writing task than for the specific writing task; whereas, the native Chinese language group used a similar number of words for both the general writing task and the specific writing task.

2. However, in the syntactic complexity measures, all the analytic scores, and the holistic scores, the NLxtopic interaction effects were not significant, indicating that the topic effects are parallel across the two different native-language groups.

3. The MGxtopic interaction effect was significant only in the holistic scores. This indicates that the hard-science major group received significantly higher scores for the specific writing task than for the general writing task; while the holistic scores of the

social-science major group were not significantly different in the two different writing tasks.

4. The nonsignificant MGxtopic interaction effects in the writing skill measures, except the holistic scores, indicate that different academic backgrounds do not significantly affect writing performance (in elaboration length, syntactic complexity, and analytic scores) for the two writing tasks.

5. The interaction effect of the three analytic main criteria and topic was significant. This result indicates that the content/thinking scores is significantly higher than the style/tone scores in the general writing task; whereas, in the specific writing task, the style/tone score is significantly higher than the content/thinking scores.

6. The interaction effect of the two scoring methods and topics was significant, indicating that the essays for the general writing task are rated highly by the analytic scoring method; whereas, the essays for the specific writing task are rated highly by the holistic scoring method.

Correlations

In the investigation of the correlations between syntactic characteristics and quality ratings, the following is the major findings:

1. Of the two types of writing tasks, the amount of prediction from syntactic complexity variables is similar for both the general writing task and for the specific writing task.

2. Among all the syntactic characteristic variables, the total number of words variable is the most important predictor, and the next important predictor is the number of T- units for the general writing task, and the mean T-unit length for the specific writing task.

3. When the total number of words variable is deleted from all the syntactic characteristic variables in the multiple regression, the elaboration-length variables are still more significant predictors than the syntactic complexity variables.

4. According to the confirmatory factor analysis for the syntactic complexity and essay quality variables, the factor intercorrelation is .497, indicating that the syntactic complexity variables measure different attributes of writing skill than the essay quality variables. For the syntactic complexity factor, the factor loading of the mean T-unit length variable (.966) is higher than the factor loading of the ratio of closers variable (.736).

The investigation of the correlations between the analytic scores and the holistic scores shows the following two major findings:

1. The relationship between the analytic scores and the holistic scores varies from topic to topic. While the holistic rating for the general writing task relies heavily on the dimensions of the content/thinking (especially, the richness of information and alternatives subcriteria) and organization (especially, the unity criterion), the holistic rating for the specific writing task focuses on the dimensions of style/tone (especially, the objectivity criterion) and organization (especially, the framing criterion).

2. According to the confirmatory factor analysis for the analytic main criteria scores of the general writing task and the analytic main criteria scores of the specific writing task, the interfactor correlation is .814. This correlation is relatively high when compared with the essay reliability estimation of the previous studies, indicating that the analytic main criteria scores of the two writing tasks (general vs. specific) do not seriously change from one topic to the other as some researchers suggest.

Finally, the investigation of the correlations between the writing skill measures and the general verbal skill measures yields the following major findings:

1. The correlation between the essay quality scores and the GRE verbal scores are stable across the two writing tasks.

2. Among all the essay quality variables, the holistic scores are the most significant predictor of the GRE verbal scores, indicating that the holistic scoring method measures similar attributes of general verbal skill as does the GRE verbal test.

3. Of the analytic scores variables, the style/tone main criterion scores make the most important contribution to the prediction of the GRE verbal scores in the two writing tasks.

4. The interfactor correlation between the writing skill factor and the general verbal skill factor is very high (.885), indicating that the two factors measure almost the same skill. The reading comprehension variable (based on the GRE scores) has similar factor loading for the writing skill factor and the general verbal skill factor.

5. The factor intercorrelation between the writing skill factor and the general verbal skill factor (.898) is higher than the factor intercorrelation between the writing skill factor and the reasoning skill factor (.773) in the three factor model.

Group Differences

Following are the major findings on group differences in writing performance:

1. The two native-language groups show significant differences in every writing skill measure used in this study. When the analytic total scores and the holistic scores are adjusted for the GRE verbal scores, the group difference is still significant in the analytic total scores and the holistic scores. This implies that the native-language group difference in the essay quality scores (measured by an essay test) is significantly larger than the group difference in the general verbal skill scores (measured by objective test).

2. The two academic-major groups do not show any significant difference in the syntactic complexity and the essay quality measures. However, the two groups show significant difference in the elaboration-length measures (the total number of words and the number of words in free modifiers measures). The social-science-major group use larger number of words in their essays than the hard-science-major group.

Rating Consistency

The major findings on the rating consistency problem include the following:

1. The reading reliability of the analytic scoring is similar for the two topics (.80 for the "Space" topic and .81 for the "Farm" topic based on two readings for one task), indicating that the rating of the essays does not vary significantly from one topic to the other. The reading reliability of two readings for the two tasks is .89 for the sample of this study. These reading reliabilities are comparatively higher than those of other studies.

2. The reading reliabilities of the analytic main criteria are affected by the different types of writing tasks. An interesting fmding is that the reading reliabilities of the analytic main criteria are better for the analytic criterion judged to be of major importance to the essay than for the analytic main criterion judged to be of minor importance the the essay.

3. The scores reliability of the two readings for the two tasks is .89, indicating that writing performance does not vary from topic to topic.

4. The results of the analysis of variance support the above findings in regard to the reliabilities. The results show that the rater group main effect and the interaction effect of the rater group and topic are not significant. However, the results also indicate a significant interaction effect of the native-language group and rater group. This signif-

icant interaction effect implies that although different raters' ratings are stable across different types of topics, the rating can be affected by writer groups with a different cultural background.

Discussion and Conclusions

This section contains discussion and conclusions on the topic effects, interaction effects, and correlations in this study.

Topic Effects

The study of the topic effects showed three major results: (a) the general writing task facilitates more elaboration (longer essay), and higher quality in the analytic content/thinking main criteria than the specific writing task, (b) both the general writing task and the specific writing task facilitate almost the same degree of syntactic complexity and the same quality in the analytic organization and analytic style/tone criteria, (c) and the specific writing task facilitates higher quality in the holistic scoring method.

The most significant effect of the information types (general vs. specific) in the writing assignment was found in the elaboration-length variables. This result supports the notion that the amount of elaboration in writing is related to the information types. For example, in the "Space" topic (general information provided), writers can draw information from their own knowledge and experience on the advantages and disadvantages of space exploration; whereas, in the "Farm" topic (specific information provided), writers have to rely heavily on the specific information given in the assignment. Therefore, the results suggest that a writing task requiring writers to draw more-

upon their

previous knowledge facilitates more elaboration as measured by the total number of words and number of words in free modifiers. This result indicates that the findings of the learning and memory research on reading comprehension can be applicable to writing research. Several researchers (Anderson et al., 1978; Bransford et al., 1978; Stein & Bransford, 1979; Benton and Blohm, 1986) reported that reading comprehension tasks requiring subjects to draw on previous knowledge resulted in higher levels of text recalls than reading comprehension tasks not requiring extensive use of previous knowledge.

The nonsignificant effect of the information types on the syntactic complexity variables supports several researchers' findings that the T-unit length and the ratio of free modifiers are stable indices of syntactic complexity and maturity (Christensen, 1968; Christensen & Christensen, 1978; Hunt, 1965 1983; Wolk, 1970). Some researchers (Crowhurst & Piche, 1979; O'Donnell, 1976; Watson, 1983) reported that syntactic complexity in written composition is significantly affected by different discourse modes and different intended audience variables. However, the results of this study provide evidence that the syntactic complexity measures (T-unit length and ratio of free modifiers) are stable across different types of writing tasks within the same discourse mode and audience.

The significant effect of the information types on the analytic content/thinking scores suggests that the general writing task obligates students to produce more extensive information, more explicit judgments, and more alternative views of thinking than the specific writing task. The nonsignificant effects of the information types on the analytic organization and style/tone scores suggest that the different information types do not significantly affect writing performance in regard to the organization of the essay and in regard to the style and manner that matches the conventions to academic English. This

result indicates that the latter two analytic criteria are not sensitive to the information types provided in the assignment.

Why do the information types significantly affect the analytic content/thinking scores but not the analytic organization scores and the style/tone scores? One of the reasons is that the content/thinking criterion focuses on declarative knowledge (knowledge of facts or ideas); whereas, the organization and style/tone criteria focus on procedural knowledge (automatized knowledge of how to do something). Declarative knowledge on the content of the essay is sensitive to the content information provided in the assignment. But procedural knowledge on the organization and style of the essay is not sensitive to the content information provided in the assignment.

The significant topic effect on the holistic scores contrasts with the topic effect on the analytic scores. The holistic scores for the specific writing task is significantly higher than the general writing task; whereas, the analytic scores for the specific writing task are lower than the general writing task. One of the problems in interpreting the above results has to do with the characteristics of the holistic scoring method. The major function of the holistic scoring method is to separate the better performers from the poor performers by rank ordering the essays; hence, the holistic scores provides little information about the "intrinsic" quality (Hirsh, 1977, p. 189) of an individual student's writing.

However, one of the reasons for this difference can be explained by the results of the investigation on the relationship between the holistic scores and the syntactic characteristic variables and the relationship between the holistic scores and the analytic subcriteria scores. According to the results, the holistic scoring focuses less on essay length variables compared to the analytic scoring, although the essays on the general writing task are significantly longer than the essays on the specific writing task. This

means that holistic scoring focuses less on the number of ideas (more extensive ideas and alternative views of thinking) than the analytic scoring. The holistic scores correlate more highly with the analytic organization subcriteria (the unity subcriteria for the general writing task and the framing subcriteria for the specific writing task) than the analytic content/thinking subcriteria. In addition, the holistic scoring method used in this study considered mechanical errors of students' essays; whereas, the analytic scoring method used in this study did not deal with the mechanical errors.

Interaction Effects

The principal results of the interaction effects suggest the following two major points: (a) the interaction of the native-language groups and the topics is significant in the elaboration-length variables, and (b) the interaction of the major groups and the topics is significant in the holistic scores.

The nonsignificant interaction of the native-language groups and the topics in the writing skill measures (except the elaboration-length measures) suggests that the two types of writing tasks affect equally the two native-language groups' writing performance in essay quality and syntactic complexity. The significant interaction of the native- language groups and the topics in the elaboration-length variables suggests that the topic effect in elaboration is stronger for the native-English group (fluent writers) than for the native-Chinese group (less fluent writers). This may be due to the constraints that writers have on the process of their writing. According to Bruce et al (1983), the major processes of writing include idea production, text production, and editing. All of these processes contribute to creating a text that satisfies several constraints such as appropriate wording, good sentence forms, paragraph form, and text form. Therefore, less

fluent writers have more severe constraints on their writing processes than fluent writers. For this reason, the less fluent writers have difficulty in producing longer essays in a limited span of time regardless of the number of ideas generated from their long-term memory or the information provided in the assignment.

The significant interaction of the academic-major group and the topics on the holistic scores suggest that the holistic scores can be affected by the students' academic background, interest, and experiences on the given topic types. On the basis of a survey of academic writing tasks, Bridgeman and Carlson (1984) reported that different academic disciplines emphasize different dimensions of writing skill. Accoiding to their report, a topic type such as describe and interpret a graph or chart is emphasized in the engineering and science departments; whereas, a topic type such as compare and contrast plus take position is emphasized in the social sciences and English departments. However, it is not clear why the significant interaction of the major groups and the topics appears only in the holistic scores. It is due to the characteristics of the holistic scoring method. As discussed in the previous section, the holistic scoring is based on a general impression of the essay, and the general (extrinsic) impression relies heavily on elements such as framing and unity. The hard-science group students' deficiency on the framing or unity for the general writing task affects significantly the rater's general impression or extrinsic judgment of the essays.

Correlations

The major fmdings on correlations between different measures of writing skill show the following three major points: (a) the correlation between the essay quality variables and the syntactic characteristic variables are stable across topics, (b) the correlations

between the analytic scores and the holistic scores vary from topic to topic, (c) and the correlations between the essay quality scores and the GRE verbal scores are almost stable across the two different types of topics.

The similar correlations between the syntactic complexity variables and the essay quality variables in both the general writing task and the specific writing task suggest that the quality ratings on the general writing task are affected equally by the syntactic complexity of the essay as the quality ratings on the specific writing task. In addition, the elaboration length variables correlate more highly with the essay quality variables than do the syntactic complexity variables. This result is consistent with other researchers' findings that the essay quality ratings focus more on the number of ideas generated in the essay rather than on the mature style of the syntax (Breland, 1983; Grobe, 1981; Hendrickson, 1980; Stewart & Grobe, 1979). This result can be explained by the fact that the syntactic complexity variables measure only the syntax level skill, while the elaboration-length variables reflect the content and structure of the overall essay. The modest degree of the estimated true correlation between the essay quality factor and the syntactic complexity factor supports the above explanation.

The different correlations between the analytic scores and the holistic scores across the two topics suggest that the holistic scoring method emphasizes different elements of writing skill for different types of topics. The holistic rating for the general writing task correlates highly with the analytic content/thinking main criterion score, whereas the holistic rating for the specific writing task correlates highly with the analytic style/tone main criterion scores. This result is further supported by the reading reliabilities of the analytic main criteria. The reading reliability of the analytic content/thinking main criterion is highest for the general writing task; while that of the analytic style/tone main

criterion is highest for the specific writing task. It was reported that the reading reliability was generally better for the scoring criterion judged to be of major importance to the essay than for the scoring criterion judged to of minor importance (Linn et al., 1972). Therefore, the different reading reliabilities for the analytic main criteria imply that the content/thinking dimension is the most important criterion for the general writing task, and the style/tone dimension is the most important criterion for the specific writing task. However, the high degree of the estimated true correlation between the analytic scores for the general writing task and the analytic scores for the specific writing task suggests that the analytic scores do not seriously change from one topic to the other as other researchers suggested (Odell et al., 1978; Breland, 1983; Meredith & Williams, 1984).

The stable correlations between the writing skill measures and the GRE verbal scores suggest that the writing skill variables measure consistently certain degrees of general verbal skill across the two different topics. The correlation between the holistic scores and GRE verbal scores is the highest, and the correlation between the analytic style/tone main criterion scores and the GRE verbal scores is next highest, suggesting that the holistic rating and the analytic style/tone criterion measure almost the same skill that the GRE verbal test measures. In addition, the estimated true correlation between the essay quality scores and the GRE verbal scores is higher than that of the essay quality scores and the GRE reasoning scores. This suggests that writing skill depends more heavily on the general verbal skill factor than it does on the reasoning skill factor.

Implications

The results of this study have several implications for theory, research, and assess- went.

For theory, the results of this study provide some information for the relationship between the writing task and the text produced. Current models of writing emphasize the role of the writing assignment (or task) in the writing processes and production. However, the process models of writing do not handle in sufficient detail the effect of the writing task on the writing processes and production for the models to be considered comprehensive models of writing. These models do not specify the relative importance of different kinds of information given in the assignment, and the models do not indicate how the influence of the information type given in the assignment varies with different groups of writers. Regarding these problems in the process models of writing, the results of this study make clear the link between the information types given in the writing assignment and the text produced.

In the field of writing theory, it has been generally accepted that the planning process plays a major role in writing (Flower and Hayes, 1984). According to Kirby (1984), the planning process consists of three main components: metacognition (decision making), selective attention (working memory), and strategies. For the three components, information is selected from either the writer's long-term memory or the information given in the writing assignment. The results of this study imply that the information selection process can be affected differently by the type of information given in the writing assignment. More importantly, the results of this study suggest that the ideas, selected more heavily from the writer's long-term memory facilitates longer elaboration, more extensive ideas, and more alternative views of thinking than when the ideas depend upon the information given in the writing assign-

ment. In addition, the results of this study provide an example of how the different in- formation types affect the writing performance of different writer groups.

For research (especially for research on the writing assignment), the results of this study add new information for the effects of the information type (general vs. specific) provided in the assignment on writing performance in the elaboration length, syntactic complexity, and essay quality. Several studies have been concerned with how the elements of the writing assignment affect student writing performance. The interest of the writing assignment study is derived not only from the assessment purpose but also from the instructional purpose (Hillocks, 1986; Odell et al., 1978). Those who are concerned with writing assessment wish to insure that a given assignment will elicit the writer's best performance, and those who are concerned with instruction wish to know how particular characteristics of an assignment that consistently produce better writing. Most of the writing assignment studies have examined the effects of the rhetorical elements such as discourse mode, discourse purpose, and audience.

More recently, some researchers have examined the effects of the information-load levels and the task-complexity levels on student writing performance. However, the type of information in the assignment has not been well investigated. The types of information or question type (in regard to general versus specific) has proved to be a significant role in memory as well as in reading comprehension in the area of reading comprehension research. In addition, the writing assignment studies have usually used a single measure of writing skill. However, writing skill is very complex, and the assignment effect may not be simple. Therefore, the single measure of writing skill may provide only limited information about the assignment effect.

For the above two reasons, this study focused on the two types of information (general vs. specific) provided in the assignment, and used several different kinds of writing skill measures. Thus, the results of this study shed light on two issues of writing assignment research: (a) a system for classification of the writing assignment variables, and (b) a system for quantifying ratings of essay quality. In regard to the above two issues, it has been suggested that the classification based on the traditional rhetorical elements (such as discourse modes and purposes) has not worked (Hoetker, 1982a), and that the holistic scores are insufficient to identify the effects of the different writing aSsignments on writing performance (Ruth and Murphy, 1984). The results of this study support these suggestions. In addition, for the studies of the writing assignment influences for students with different cultural or academic background, the results of this study specify the topic effects on several different aspects of writing skill across different writer groups. Finally, the research design (split-plot factorial design) used in this study was proved to be appropriate for the analyses of varied samples of writing from the same student.

For writing assessment, the results of this study clarify in sufficient detail not only the effects of the different information types provided in the assignment on student writing performance but also the relationships among the several writing skill measures. As shown in the discussion section, the results of this study suggest several points about topic effects, the interaction effects of the topics and the writer groups, the interaction effects of the topics and scoring method, and the correlations between different measures of writing skill. On the basis of these results, the following implications appear for the writing assessment. First, in a large scale assessment, the holistic scoring can be used economically for a selection purpose. However, a single writing assignment will cause a measurement problem because the holistic scores vary significantly from topic to topic. When the analytic total scores are used for a selection purpose, a single assignment is

enough because the analytic total scores remain stable across different topics. Second, in a school writing program, the analytic scoring method can be used more appropriately for diagnostic purposes. The analytic main criteria and subcriteria scores can provide diagnostic information about the strengths and weaknesses of student writing.

Much more research is needed to clarify the topic effects on student writing performance for a reliable and valid assessment of writing skill.

First, more research is needed to define the levels of writing skill. The analytic main criteria and subcriteria used in this study may be appropriate for the assessment of college-level students' writing performance; however, it is uncertain whether these analytic criteria are appropriate for the assessment of elementary level or high-school level students' writing performance.

Second, more studies on the topic effects of several different samples of writing from the same students are needed. This study used only two kinds of samples of writing from the same student in regard to the type of information (general vs. specific). Further research is needed to clarify the combined effects of the information type and the task complexity levels using several different kinds of samples from the same students.

Third, further research is needed to clarify the topic effects on the characteristics of the elaboration. This study used simple quantitative measures (such as total number of words, total number of T-Units, and total number of words in free modifiers) for elaboration length. However, these measures cannot clearly provide information about the quality of the essay structure. The characteristics of the elaboration can be measured by using several researchers' discourse analysis systems (Frederiksen, 1975, 1977; Kintsch & van Dijk, 1978; Meyer, 1975). The studies based on these discourse analysis systems may clarify not only the topic effects on the characteristics of elaboration but also the relationships between the essay structure and the essay quality scores.

REFERENCES

Anderson, J. R., & Reader, L. M. (1978). An elaborative processing. In L. S. Cermak & F. I. M. Craik (Eds.), ***Levels of processing and human memory*** (pp. 385-403). Hillsdale, NJ: Erlbaum.

Applebee, A. N. (1984). Writing and reasoning. ***Review of Educational Research, 54,*** 577-596.

Bain, A.(1890/1980). ***English composition and rhetoric: A manual*** (rev. ed.). New York: Appleton.

Bates, P.T. (1985). Writing performance and its relationship to the writing attitudes, topic knowledge, and writing goals of college freshmen (Doctoral dissertation, University of Texas at Austin, 1984). ***Dissertation Abstracts International, 46,*** 366A-367A.

Bauer, B. A. (1982). ***The reliabilities and the cost efficiency of three methods of assessment of writing ability: An empirical inquiry.*** Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Beaugrande, R. de. (1984). ***Text production.*** Norwood, NJ: Ablex.

Benton, S. L., & Blohm, P. J. (1986). Effect of question type and position on measures of conceptual elaboration in writing. ***Research in the Teaching of English, 20.*** 98-108.

Bereiter, C. (1980). Development in writing. In L. W. Gregg & E. R. Steinberg(Eds.),***Cognitive processes in writing*** (PP 73-93). Hillsdale, NJ: Erlbaum.

Bloom, B. S. (Ed.) (1956). ***Taxonomy of educational objectives: Handbook I: Cognitive domain.*** New York: McKay.

Bobrow, S. A., & Bower, G. H. (1969). Comprehension and recall of sentences. ***Journal of Experimental Psychology, 80,*** 455-461.

Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). ***Research in written composition.*** Champaign, IL: National Council of Teachers of English.

Bransford, J. **D.,** Franks, J. J., Morris, C. D., & Stein, B. S. (1978). Some general constraints on learning and memory research. In L. S. Cermak & P. I. M. Craik (Eds.), ***Levels of processing and human memory*** (pp. 331-354). Hillsdale, NJ: Erlbaum.

Breland, H. M. (1983). ***The direct assessment of writing skill: A measurement review*** (Report No. 83-6). New York: College Entrance Examination Board.

Breland, H. M., & Gaynor, J. L. (1979). A comparison of direct and indirect assessments of writing skill. ***Journal of Educational Measurement, 16,*** 119-128.

Breland, H. M., & Griswold, P. A. (1981). ***Group comparisons for basic skill measures***

(Report No. 81-6). New York: College Entrance Examination Board.

Breland, H. M., & Jones, R. J. (1982). ***Perception of writing skill*** (Report No. 82-4).

New York: College Entrance Examination Board.

Brewer, W. F. (1981). Literary theory, rhetoric, and stylistics: Implication for psychology. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), ***Theoretical issue in reading comprehension*** (pp. 221-239). Hillsdale, NJ: Erlbaum.

Bridgeman, B., & Carlson, S. B. (1984). Survey of academic writing tasks. ***Written Communication, 1,*** 247-280.

Britton, J., Burges, T., Martin, N., Mcleod, A., & Rosen, H. (1975). ***The development of writing abilities (11-18).*** Urbana, IL: National Council of Teachers of English. Brooks, C., & Warren, R. P. (1972). ***Modern rhetoric*** (3rd. ed.). New York: Harcourt Brace.

Brossell, G. (1983). Rhetorical specification in essay examination topics. ***College English, 45,*** 165-173.

Bruce, B., Collins, A., Rubin, A. D., & Gentner, D. (1983). ***Three perspectives on writing*** (Reading Education Report No. 41). Champaign: University of Illinois, Center for Study of Reading.

Burleson, B. R., & Rowan, K. E. (1985). Are social cognitive ability and narrative writing skill related? A response to Rubin et al. ***Written Communication, 2,*** 25-53.

Carlman, N. (1985). The effects of scoring method, topic, and mode on grade 12 students' writing scores (Doctoral dissertation, Simon Fraser University, Canada, 1984). ***Dissertation Abstracts International, 46,*** 367A.

Christensen, F. (1968). ***The Christensen rhetoric program.*** New York: Harper. Christensen, F., & Christensen, B. (1978). ***Notes toward a new rhetoric.*** New York: Harper.

Clemson, E. (1978). ***A study of basic skills assessment: Direct and indirect measures of writing ability.*** Princeton , NJ: Educational Testing Service.

Coffman, W. E. (1966). On the validity of essay test of achievement. ***Journal of Educational Measurement, 3,*** 151-156.

Coffman, W. E. (1971a). Essay examination. In R.L. Thorndike (Ed.), ***Educational measurement*** (2nd ed.) (pp. 270-302). Washington D.C.: American Council of Education.

Coffman, W. E. (1971b). On the reliability of ratings of essay examinations in English. ***Research in the Teaching of English, 5,*** 24-36.

Combs, W. E. (1976). Further effects of sentence-combining on writing ability. ***Research in the Teaching of English, 10,*** 137-149.

Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper & L. Odell (Eds.), ***Evaluating Writing: Describing, measuring, Judging*** (pp. 3-31). Urbana, IL: National Council of Teachers of English.

Cooper, C. R. (1984). Studying the writing abilities of a university freshman class: Strategies from a case study. In L. S. Flower & J. R. Hayes (Eds.), ***New direction in composition research*** (pp. 19-52). New York: Guilford Press.

Coward, A. F. (1952). A comparison of two methods of grading English compositions. ***Journal of Educational Research, 46,*** 81-93.

Crowhurst, M. (1980). Syntactic complexity and teachers' quality ratings narrations and arguments. ***Research in the Teaching of English, 14,*** 223-231.

Crowhurst, M., & Piche, G.L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. ***Research in the Teaching of English, 13,*** 101-109.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. ***Psychometrika, 40,*** 137- 149.

D'Angelo, F. J. (1976). Modes of discourse. In G. Tate (Ed.), ***Teaching composition: 10 bibliographical essays*** (pp 111-135). Forst Worth: Texas Christian University press. D'Angelo, F. J. (1980). ***Process and thought in composition*** (2nd ed.). Cambridge, MA: Winthrop Publishers, Inc.

Diederich, P. B. (1966). How to measure growth in writing ability. ***English Journal, 55,*** 435-449.

Diederich, P. B. (1974). ***Measuring growth in English.*** Urbana, IL: National Council of Teachers of English.

Diederich, P. B., French, J. W., & Carton, S. T. (1961). ***Factors in the judgments of writing ability*** (Research Bulletin No. 61-15). Princeton, NJ: Educational Testing Service.

Dixon, E. (1970). ***Index of syntactic maturity*** (Dixon-Hunt- Christensen). (ERIC Document Reproduction Service No. ED 091 748)

Emig, J. (1971). ***The composing process of twelth graders.*** Urbana, IL: National Council of Teachers of english.

Faigley, L. (1979). The influence of generative rhetoric on the syntactic maturity and writing effectiveness of college freshmen. ***Research in the Teaching of English, 13,*** 197-206.

Flower, L. S., & Hayes, J. R. (1977). Problem-solving strategies and the writing process. ***College English, 39,*** 449-461.

Flower, L. S., & Hayes, J. R. (1982). Plans that guide the composing process, In C. H. Frederiksen & J. F. Dominic (Eds.), ***Writing: The nature, development, and teaching of written communication: Vol. 2: Writing: Process, development and communication*** (pp 39-58). Hillsdale, NJ: Erlbaum.

Flower, L. S., & Hayes, J. R. (1984). Images, Plans, and prose. ***Written Communication, 1,*** 120-160.

Frederiksen, C. H. (1975). Acquisition of semantic informationfrom discourse: Effects of repeated exposures. ***Journal of Verbal Learning and Verbal Behavior, 14,*** 158-169.

Frederiksen, C. H. (1977). Semantic processing units in understanding text. In R. 0. Freedle (Ed.), ***Discourse production and comprehension: Vol 1: Discourse processes*** (pp. 57-87). Norwood, NJ: Ablex.

Freedman, S. W., & Calfee, R. C. (1983). Holistic assessment of writing: Experimental design and cognitive theory. In P. Mosenthal, L. Tamor, & S. A. Walmsley (Eds.), ***Research on writing*** (pp. 75-98). New York: Longman.

Gebhard, A. 0. (1978). Writing quality and syntax. ***Research in the Teaching of English, 12,*** 211-231.

Glover, J. A., Plake, B. S., Roberts, B., Zimmer, J. W., & Palmere, M. (1981). Distinctiveness of encoding: The effects of paraphrasing and drawing inferences on memory from prose. *Journal of Educational Psychology*, 73, 736- 744.

Glover, J. A., Plake, B. S., & Zimmer, J. W. (1982). Distinctiveness of encoding for learning tasks: Recall of text materials. *Journal of Educational Psychology, 74,* 189-198.

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *Measurement of writing ability.* New York: College Entrance Examination Board.

Gould, J. D. (1978). An experimental study of writing, dictating, and speaking. In J. Requin (Ed.), *Attention and performance VII* (pp. 299-320). Hillsdale, NJ: Erlbaum.

Gould, J. D. (1980). Experiments on composing letters: Some facts, some myths, and some observation. In L. W. Gregg & E. R. Stein (Eds.), *Cognitive process in writing* (pp. 97-128). Hillsdale, NJ: Erlbaum.

Greenberg, K. L. (1981). *The effect of variations in essay questions on the writing performance of CUNY freshmen* (Research Monograph Series No. 1). New York: City University of New York. (Eric Document Reproduction Service No. ED 236 266)

Greenberg, K. L. (1986). The development and validation of the TOFEL writing test: A discourse of TOFEL research reports 15 and 19. *TESOL Quarterly, 20,* 531-544.

Grobe, C. H. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 15,* 75-85.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing process. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive process in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hayes, J. R., & Flower, L. S. (1983). Uncovering cognitive processing in writing: An introduction to protocol analysis. In P. Mosenthal, L. Tamon, & S. Walmsley (Eds.), ***Research on writing: principle and methods*** (pp. 206-220). New York: Longman.

Harmon, H. H. (1976). ***Modern factor analysis*** (3rd ed.). Chicago, IL: University of Chicago Press.

Hendrickson, L. (1980). Procedures and results of an evaluation of writing. ***Educational Evaluation and Policy Analysis, 2,*** 19-30.

Hillocks, G. Jr. (1986). ***Research on written composition.*** Urbana, IL: National Council of Teachers of english.

Hirsch, E. D. Jr. (1977). ***The philosophy of composition.*** Chicago, IL: The University of Chicago Press.

Hirsch, E. D. Jr., & Harrington, D. P. (1981). Measuring the communicative effectiveness of prose. In C. H. Frederiksen & J. F. Dominic (Eds.), ***Writing: The nature, development, and teaching of written communication: Vol. 2. Writing: Process, development and communication*** (pp. 189-208). Hillsdale, NJ: Erlbaum.

Hoetker, J. (1982a). Essay examination topics and students' writing. ***College Composition and Communication, 33,*** 377-391.

Hoetker, J. (1982b). ***Effects of essay topics on student writing: A review of literature.*** (Eric Document Reproduction Service No. ED 217 486)

Hunt, K. W. (1965). ***Grammatical structures written at three grade levels.*** Champaign, IL: National Council of Teachers of English.

Hunt, K. W. (1970). ***Syntactic maturity in school children and adults*** (Society for Research in Child Development Monographs No. 134). Chicago, IL: The University of Chicago Press.

Hunt, K. W. (1977). Early blooming and late blooming syntactic structures. In C. R. Cooper & L. Odell (Eds.), ***Evaluating writing: Describing, measuring, judging*** (pp. 91-104). Urbana, IL: National Council of Teachers of English.

Hunt, K. W. (1983). Sentence combining and the teaching of English. In M. Martlew (Ed.), ***The psychology of written language*** (pp. 99-125). New York: John Willey and Sons.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). ***Testing ESL composition: A practical approach.*** Rouley, MA: Newbury House Publishers.

Joreskog, K. G., & Sorbom, D. (1979). ***Advances in factor analysis and structural equation models.*** Cambridge, MA: Abt Books.

Joreskog, K. G., & Sorbom, D. (1984). ***LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood*** (3rd ed.). Mooresville, IN: Scientific Software Inc.

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixation to comprehension. ***Psychological Review, 87,*** 329-354.

Kaiser, H. F., & Cerny, B. A. (1979). Factor analysis of the image correlation matrix. ***Educational and Psychological Measurement, 39,*** 711-714.

Kinneavy, J. L. (1971). ***A theory of discourse.*** Englewood Cliffs, NJ: Prentice Hall Inc. Kinneavy, J. L. (1980). A plurastic synthesis of four contemporary models for teaching composition. In A. Freedman & I. Pringle (Eds.), ***Reinventing the rhetorical tradition*** (pp. 37-52). Conway, AR: L & S Books.

Kintch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. ***Psychological Review, 85,*** 363-394.

Kirby, J. R. (1984). Educational roles of cognitive plans and strategies. In J. R. Kirby (Ed.) ***Cognitive strategies and educational performance*** (pp. 51-88). New York: Academic Press.

Kirk, R. E. (1982). ***Experimental design: Procedures for the behavioral sciences.*** Belmont, CA: Brooks/Cole Publishing.

Kroll, B. M. (1985). Social cognitive ability and writing performance: How they are related? ***Written Communication,2,*** 293-305.

Linn, R. L., Klein, S. P., & Hart, F. M. (1972). The nature and correlates of law school essay grades. ***Educational and Psychological Measurement,*** 32, 267-279.

Martlew, M. (1983). Problems and difficulties: Cognitive and communicative aspects of writing development. In M. Martlew (Ed.), ***The psychology of written language*** (pp. 295- 333). New York: John Wiley and Sons.

Meredith, V. H., & Williams, P. L. (1984). Issues in direct writing assessment: Problem identification and control. ***Educational Measurement: Issues and practice, 3,*** 11-35.

Meyer, B. J. F. (1975). ***The organization of prose and its effects on memory.*** Amsterdam. North Holland Publishing Co.

Meyer, B. J. F., & Freedle, R. 0. (1984). Effect of discourse type on recall. ***American Educational Journal, 21,*** 121-143.

Mitchell, D. C. (1982). ***The process of reading.*** New York: John Wiley and Sons. Moffett, J. (1968). ***Teaching the universe of discourse.*** Boston, MA: Houghton Mifflin. Mosenthal, P. (1983). Defining classroom writing competences: A paradigmatic perspective. ***Review of Educational Research,*** 53, 217-251.

Mullis, I. (1984). Scoring direct writing assessment: What are the alternatives? ***EducationalMeasurement: Issues and Practice,*** 3, 16-18.

Myers, M. (1980). ***A procedures for writing assessment and holistic scoring.*** Urbana, IL: National Council of Teachers of English.

Nold, E. W., & Freedman, S. W. (1977). An analysis of readers' responses to essays. ***Research in the Teaching of English, II,*** 164-174.

Odell, L. (1979). Teachers of composition and needed research in discourse theory. ***College Composition and communication, 30,*** 39-44.

Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.) ***The nature and measurement of competency in English*** (pp. 139-154). Urbana, IL: National Council of Teachers of English.

Odell, L., Cooper, C. R., & Courts, C. (1978). Discourse theory: Implication for research in composing. In C. R. Cooper & L. Odell (Eds.), ***Research on composing*** (pp. 1-12). Urbana, IL: National Council of Teachers of English.

O'Donnell, R. C. (1976). A critique of some indices of syntactic maturity. ***Research in the Teaching of English, 10,*** 31-38.

O'Donnell, R. C. (1977). Language development: Kindegarten through grade twelve: A review. ***Research in the Teaching of English, 11,*** 49-53.

O'Hare, F. (1973). ***Sentence combining: Improving student writing without formal grammar instruction.*** Urbana, IL: National Council of Teachers of English.

Pianko, S. (1979). A description of the composing processes of college freshman writers. ***Research in the Teaching of English, 13,*** 5-22.

Prater, D. (1982). ***The effect of writing task on mastery classification of high school students.*** Paper presented at the annual meeting of the Southwest Educational Research Association (Austin, TX, February). (Eric Document Reproduction Service No. ED 215 367)

Prater, D., & Padia, W. (1983). Effects of modes of discourse on writing performance in grade four and six. ***Research in the Teaching of English, 17,*** 127-134.

Purves, A. C. (1984). In search of an internationally-valid scheme for scoring composition. ***College Composition and Communication, 35,*** 426-438.

Purves, A. C. (1985). ***Framework for scoring: GRE/TOEFL.*** Unpublished manuscript, University of Illinois at Urbana- Champaign, Curriculum Laboratory.

Purves, A. C., Soter, A., Takala, S., & Vahapassi, A. (1984). Towards a domain- referenced system for classifying composition assignments. ***Research in the Teaching of English, 18,*** 385-409.

Quellmalz, E., & Capell, F. (1979). ***Defining writing domains: Effects of discourse and response mode.*** California University, Los Angeles, Center for the Styudy of Evaluation. (Eric Document Reproduction Service No. ED 212 661)

Quellmalz, E., Capell, F., & Chou, C. P. (1982). Effects of discourse and response mode on the measurement of writing competence. ***Journal of Educational Measurement, 19,*** 241- 258.

Rafoth, B. A. (1985). Audience adaptation in the essays of proficient and non proficient freshmen writers. ***Research in the Teaching of English, 19,*** 237-253.

Reed, W. M., Burton, J. K., & Kelly, P. P. (1985). The effect of writing ability and mode of discourse on cognitive capacity engagement. ***Research in the Teaching of English, 19,*** 283-297.

Remodino, C. (1959). A factorial analysis of the evaluation of scholastic compositions in the mother tongue. ***British Journal of Educational Psychology, 29,*** 242-251.

Rock, D. A., Werts, C., & Grandy, J. (1982). ***Construct validity of the GRE aptitude test scores across populations: An empirical confirmatory study*** (GRE Board Professional Report No. 78-1P). Princeton, NJ: Educational Testing Service.

Rubin, D. L., & Piche, G. L. (1979). Development in syntactic and strategic aspects of audience adaptation skills in written persuasive communication. ***Research in the Teaching of English, 13,*** 293-316.

Rubin, D. L., Piche, G. L., Michlin, M. L. & Johnson, F. L. (1984). Social cognitive ability as a predictor of the quality of fourth-graders' written narratives. In R. Beach & L. Bridwell (Eds.), ***New Directions in composition research*** (pp. 297-308). New York: Guilford Press.

Rubin, D. L., & Rafoth, B. A. (1986). Social cognitive ability as a predictor of the quality of expository and persuasive writing among college freshmen. ***Research in the Teaching of English, 19,*** 237-253.

Ruddell, R. B., & Speaker, R. B. (1986). The interactive reading process: A model. In H. Singer & R. B. Ruddell (Eds.), ***Theoretical models and processes of reading*** (3rd ed.) (pp. 751-793). Newark, DE: International Reading Association.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), ***Attention and performance VI*** (pp. 573-603). Hillsdale, NJ: Erlbaum.

Slamecka, N. J., & Graf, P. (1978). The general effect: Delineation of a phenomenon. ***Journal of Experimental Psychology, 4,*** 592-604.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. ***College Composition and Communication, 31,*** 378-388.

Soter, A. (1985). ***GREI TOEFL scoring criteria.*** Unpublished manuscript, University of Illinois at Urbana-Champaign, Curriculum Laboratory.

Spandal, V., & Stiggins, R. J. (1981). ***Direct measures of writing skill: Issues and Application.*** Portland, OR: Northwest Regional Educational Laboratory.

Steele, A. (1979). ***The assessment of writing proficiency via qualitative ratings of writing samples.*** Paper presented at the annual meeting of the National Council on Measurement in Education (Sanfransisco, CA, April). (Eric Document Reproduction Service No. ED 175 944)

Stein, B. B., & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. ***Journal of Verbal Leraning and Verbal Behavior, 18,*** 769-777.

Stewart, M. F. (1978). Syntactic maturity from high school to university: A first look. ***Research in the Teaching of English, 12,*** 37-46.

Stewart, M. F., & Grobe, C. H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality rating. ***Research in the Teaching of English, 13,*** 207-215.

Taylor, I., & Taylor, M. M. (1983). ***The psychology of reading.*** New York: Academic Press.

Tibbetts, A. M., & Tibbetts, C. (1984). ***Strategies of rhetoric with handbook*** (4th ed.). Glenview, IL: Scott, Foresman and Company.

Tierney, R. J., & Pearson, P. D. (1983). ***Toward a composing model of reading*** (Reading Education Report No. 43). Champaign: University of Illinois, Center for the Study of Reading.

Thorndike, R. L. (1982). ***Applied Psychometrics.*** Boston, MA: Houghton Mifflin Company.

Vahapassi, A. (1982). On the specification of the domain of school writing. ***Evaluation in Education: An International Review Series,*** 5, 265-289.

van Dijk, T. A. (Ed.) (1985). ***Handbook of discourse analysis: Vol. 2: Dimensions of discourse.*** New York: Academic Press.

Veal, L. R., & Tillman, M. (1971). Mode of discourse variation in the evaluation of childrens' writing. ***Research in the Teaching of English, 5,*** 37-45.

Watson, C. (1983). Syntactic changes: Writing development and the rhetorical context. In M. Martlew (Ed.), ***The Psychology of written language*** (pp. 127-139). New York: John Willey and Sons.

Werts, C. E., Breland, H. M., Grandy, J., & Rock, D. R. (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors. ***Educational and Psychological Measurement, 40,*** 19-29.

Witte, S. P., & Davis, A. S. (1980). The stability of T-Unit length: A preliminary investigation. ***Research in the Teaching of English, 14,*** 5-18.

Witte, S. P., & Davis, A. S. (1982). The stability of T-Unit length in the written discourse of college freshmen: A second study. ***Research in the Teaching of English, 16,*** 71-84.

Wolk, A. (1970). The relative importance of the final free modifier: A quantitative analysis. ***Research in the Teaching of English, 4,*** 59-68.

APPENDIX A

ANALYSIS OF VARIANCE SUMMARY TABLES

Table A-1

Holistic Score as Dependant Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	319.04	284.70	.0001
MG	1	3.66	3.26	.0741
NLxMG	1	.69	.61	.4351
BL(NLxMG)	92	1.12		
Topic	1	1.78	5.66	.0194
NLxTopic	1	.29	.93	.3372
MGxTopic	1	1.60	5.07	.0267
NLxMGxTopic	1	1.25	3.98	.0491
TopicxBL(NLxMG)	92	1.31		
Total	191	2.41		

Note. In the SPF 22.2 design, the NL (native language group), MG (major group), and Topic ("general" vs. "specific") were assumed to be fixed factors, and BL (students) were assumed to be a random factor.

Table A-2

Analytic Total Score as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	533.97	203.87	.0001
MG	1	5.91	2.25	.1367
NLxMG	1	.31	.12	.7316
BL(NLxMG)	92	2.62		
Topic	1	.47	.50	.4824
NLxTopic	1	.08	.09	.7704
MGxTopic	1	.81	.85	.3579
NLxMGxTopic	1	.15	.16	.6912
TopicxBL(NLxMG)	92	.95		
Total	191	4.56		

Table A-3

Analytic Content/Thinking Score as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	55.10	176.15	.0001
MG	1	.63	2.01	.1592
NLxMG	1	.04	.12	.7270
BL(NLxMG)	92	.31		
Topic	1	.61	4.26	.0419
NLxTopic	1	.03	.19	.6650
MGxTopic	1	.20	1.36	.2460
NLxMGxTopic	1	.12	.80	.3726
TopicxBL(NLxMG)	92	.14		
Total	191	.52		

Table A-4

Analytic Organization Score as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	63.21	177.24	.0001
MG	1	.86	2.41	.1244
NLxMG	1	.11	.30	.5879
BL(NLxMG)	92	.36		
Topic	1	.12	.78	.3790
NLxTopic	1	.04	.27	.6055
MGxTopic	1	.02	.16	.6928
NLxMGxTopic	1	.00	.02	.8792
TopicxBL(NLxMG)	92	.16		
Total	191	.58		

Table A-5

Analytic Style/Tone Score as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	59.82	199.49	.0001
MG	1	.50	1.68	.1982
NLxMG	1	.00	.00	.9476
BL(NLxMG)	92	.30		
Topic	1	.20	1.50	.2230
NLxTopic	1	.11	.80	.3730
MGxTopic	1	.09	.69	.4093
NLxMGxTopic	1	.01	.09	.7661
TopicxBL(NLxMG)	92	.13		
Total	191	.53		

Table A-6

Total Number of Words as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	666701	98.58	.0001
MG	1	35970	5.32	.0223
NLxMG	1	4941	.73	.3949
BL(NLxMG)	92	6763		
Topic	1	54945	29.86	.0001
NLxTopic	1	13770	7.47	.0075
MGxTopic	1	2852	1.55	.2167
NLxMGxTopic	1	910	.49	.4841
TopicxBL(NLxMG)	92	1844		
Total	191	8230		

Table A-7

Total Number of T-units as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	360.25	8.51	.0044
MG	1	19.38	.46	.5004
NLxMG	1	22.00	.52	.4728
BL(NLxMG)	92	42.35		
Topic	1	125.13	9.74	.0024
NLxTopic	1	18.13	1.41	.2378
MGxTopic	1	2.76	.21	.6443
NLxMGxTopic	1	5.01	.39	.5340
TopicxBL(NLxMG)	92	12.84		
Total	191	29.48		

Table A-8

Number of Words in Free Modifiers as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	81633	76.84	.0001
MG	1	4770	4.49	.0368
NLxMG	1	71	.07	.7962
BL(NLxMG)	92	1062		
Topic	1	4730	9.58	.0026
NLxTopic	1	2707	5.48	.0214
MGxTopic	1	1097	2.22	.1395
NLxMGxTopic	1	5197	10.53	.0016
TopicxBL(NLxMG)	92	494		
Total	191	1274		

Table A-9

Mean T-unit Length as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	59.82	199.49	.0001
MG	1	.50	1.68	.1982
NLxMG	1	.00	.00	.9476
BL(NLxMG)	92	.30		
Topic	1	.20	1.50	.2230
NLxTopic	1	.11	.80	.3730
MGxTopic	1	.09	.69	.4093
NLxMGxTopic	1	.01	.09	.7661
TopicxBL(NLxMG)	92	.13		
Total	191	.53		

Table A-10

Mean Free Modifier Lengh as Dependent Variable

(split-plot factorial 2NL 2MG. 2Topic design)

Source	df	MS	F value	p value
NL	1	253.99	27.95	.0001
MG	1	17.99	1.98	.1628
NLxMG	1	7.14	.79	.3777
BL(NLxMG)	92	9.09		
Topic	1	.00	.00	.9858
NLxTopic	1	3.24	.59	.4449
MGxTopic	1	.22	.04	.8420
NLxMGxTopic	1	45.82	8.32	.0049
TopicxBL(NLxMG)	92	5.51		
Total	191	8.75		

Table A-11

Analytic Main Category Score as Dependent Variable

(split-plot factorial 2NL. 2Topic 3MC design)

Source	df	MS	F value	p value
NL	1	177.99	203.06	.0001
BL(NL)	94	.88		
Topic	1	.16	.50	.4800
NLxTopic	1	.03	.09	.7700
TopicxBL(NL)	94	.31		
MC	2	1.73	36.49	.0001
NLxMC	2	.07	1.48	.2298
MCxBL(NL)	188	.05		
TopicxMC	2	.39	6.82	.0014
NLxTopicxMC	2	.08	1.29	.2764
TopicxMCxBL(NL)	188	.06		
Total	575			

Note. In the SPF 2.23 design, the MC (main category), NL, and Topic were assumed to be fixed factors.

Table A-12

Analytic Total Score of Two Rator Groups as Dependent Variable

(split-plot factorial 2NL. 2Topic 2RG design)

Source	df	MS	F value	p value
NL	1	1067.94	203.06	.0001
BL(NL)	94	5.26		
Topic	1	.94	.50	.4800
NLxTopic	1	.16	.09	.7700
TopicxBL(NL)	94	1.88		
RG	1	2.94	1.90	.1718
NLxRG	1	18.29	11.78	.0009
RGxBL(NL)	94	1.55		
TopicxRG	1	1.52	.85	.3584
NLxTopicxRG	1	.12	.07	.7970
TopicxRGxBL(NL)	94	1.79		
Total	383	5.42		

Note. In the SPF 2.22 design, the RG (rator group) was assumed to be a fixed factor.

Table A-13

Holistic Score and Transformed Analytic Score as Dependent Variable

(split-plot factorial 2NL. 2Topic 2SM design)

Source	df	MS	F value	p value
NL	1	364.55	316.03	.0001
BL(NL)	94	1.15		
Topic	1	.37	1.29	.2591
NLxTopic	1	.08	.28	.5955
TopicxBL(NL)	94	.28		
SM	1	141.22	344.19	.0001
NLxSM	1	38.04	92.70	.0001
SMxBL(NL)	94	.41		
TopicxSM	1	1.65	7.73	.0066
NLxTopicxSM	1	.23	1.09	.2993
TopicxSMxBL(NL)	94	.21		
Total	383	1.93		

Table A-14

Rater Reliability Estimate of the "Space" data.

Source	df	MS	F value	p value
BL	95	9.67	5.11	.0001
RG	1	.12	.06	.8052
Resudal	95	1.89		

Note. Reliability = (BLMS-resudalMS)/BLMS = .8043

Table A-15

Rater Reliability Estimate of the "Farm" Data

Source	df	MS	F value	p value
BL	95	8.63	5.37	.0001
RG	1	4.35	2.71	.1031
Resudal	95	1.61		

Note. Reliability = (BLMS-Resudal MS)/BLMS = .8139

Table A-16

ANCOVA Summary Table: Holistic Score as Dependent Variable and GRE Verbal Score (VL) as Covariate.

(randomized block design)

Source	df	MS	F value	p value
NL	1	13.53	24.38	.0001
BL	47	.27	.48	.9932
VL	1	3.79	6.83	.0121
Error	46	.55		

Table A-17

ANCOVA Summary Table: Analytic Total Score as Dependent Variable and GRE Verbal Score (VL) as Covariate.

(randomized block design)

Source	df	MS	F value	p value
NL	1	378.27	18.03	.0001
BL	47	15.87	.76	.8280
VL	1	129.21	6.16	.0168
Error	46	20.98		

APPENDIX B

CORRELATIONAL DATA

Table B-1

Intercorrelation of the Syntactic Characteristic Variables, the Analytic Total Scores, and the Holistic scores for the "Space" Topic

	Variables												
	TW	TU	WO	WI	WC	WF	MT	MO	MI	MC	MF	ST	SP
TW	1												
TU	.68	1											
WO	.46	.27	1										
WI	.48	.21	.17	1									
WC	.58	.14	.11	.25	1								
WF	.75	.29	.66	.49	.79	1							
MT	.44	-.31	.20	.39	.59	.59	1						
MO	.14	-.14	.87	.08	.06	.53	.32	1					
MI	.35	.02	.10	.95	.24	.43	.46	.08	1				
MC	.34	-.16	.02	.19	.90	.66	.71	.10	.23	1			
MF	.40	-.18	.49	.39	.73	.84	.78	.61	.44	82	1		
ST	.80	.44	.32	.38	.56	.64	.45	.10	.34	.37	.40	1	
SP	.75	.36	.35	.38	.54	.64	.48	.16	.33	.38	.43	.82	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 12 and Table 13. The Variable names are as follows:
TW: Total number of Words
TU: total number of T-Units
WO: total number of Words in Opners
WI: total number of Words in Interrupters
WC: total number of Words in Closers
WF: total number of Words in Free modifiers
MT: Mean T-unit length
MO: Mean Opner length (WO divided by TU)
MI: Mean Interrupter lenght (WI divided by TU)
MC: Mean Closer lenght (WC divided by TU)
MF: Mean Free modifier length (WF divided by TU)
ST: "Space" data Total analytic scores
SP: "Space" data holistic scores

Table B-2

Intercorrelation of the Syntactic Characteristic Variables, the Analytic Total Scores, and the Holistic scores for the "Farm" Topic

	Variables												
	TW	TU	WO	WI	WC	WF	MT	MO	MI	MC	MF	FT	FR
TW	1												
TU	.76	1											
WO	.42	.24	1										
WI	.42	.01	.16	1									
WC	.49	.21	.10	.35	1								
WF	.65	.25	.61	.62	.79	1							
MT	.26	-.39	.21	.59	.39	.54	1						
MO	-.11	-.36	.69	.06	-.02	.35	.46	1					
MI	.23	-.17	.03	.92	.29	.49	.66	.06	1				
MC	.15	-.20	.01	.35	.83	.64	.57	.13	.42	1			
MF	.11	-.36	.38	.55	.58	.73	.79	.62	.60	.80	1		
FT	.78	.49	.19	.38	.47	.51	.37	-.13	.31	.24	.18	1	
FR	.61	.28	.19	.48	.47	.54	.42	-.05	.40	.35	.32	.79	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 12 and Table 13. The Variable names are as follows:
TW: Total number of Words
TU: total number of T-Units
WO: total number of Words in Opners
WI: total number of Words in Interrupters
WC: total number of Words in Closers
WF: total number of Words in Free modifiers
MT: Mean T-unit length
MO: Mean Opner length (WO divided by TU)
MI: Mean Interrupter lenght (WI divided by TU)
MC: Mean Closer lenght (WC divided by TU)
MF: Mean Free modifier length (WF divided by TU)
FT: "Farm" data Total analytic scores
FR: "Farm" data holistic scores

Table B-3

Intercorrelation of the Analytic Main Criteria Scores, the Holistic Scores, and the Syntactic Complexity Variables for the "Space" Data

	Variables					
	SP	SA	SB	SC	MT	MC
Holistic Scores (SP)	1					
Content/Thinking Scores (SA)	.848	1				
Organization Scores (SB)	.841	.940	1			
Style/Tone Score (SC)	.807	.913	.919	1		
Mean T-unit Length (MT)	.475	.455	.460	.459	1	
Mean Closer Length (MC)	.378	.336	.352	.359	.711	1

Note. The above intercorrelation matrix (n=96) was used in the confirmatory factor analysis as shown in the Table 14. The analytic main criteria scores were based on the mean scores of the four raters' ratings.

Table B-4

Intercorrelation of the Analytic Subcriteria Scores and the Holistic Scores for the "Space" Data

	Variables													
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	SP
S1	1													
S2	.83	1												
S3	.83	.88	1											
S4	.75	.83	.86	1										
S5	.80	.77	.82	.81	1									
S6	.79	.77	.84	.83	.79	1								
S7	.76	.69	.72	.72	.80	.76	1							
S8	.79	.83	.80	.72	.74	.74	.77	1						
S9	.82	.71	.79	.77	.77	.75	.77	.75	1					
S10	.78	.79	.87	.78	.75	.82	.71	.77	.78	1				
S11	.82	.84	.87	.76	.77	.80	.76	.85	.77	.82	1			
S12	.71	.75	.80	.75	.70	.78	.71	.77	.70	.77	.78	1		
S13	.73	.78	.81	.69	.71	.72	.63	.80	.68	.79	.83	.74	1	
SP	.70	.79	.80	.75	.68	.74	.70	.73	.67	.81	.72	.71	.69	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 16. The Variable names are as follows:
S1: Adequacy if Information
S2: Richness of Information
S3: Relationships
S4: Inferences
S5: Synthesisn
S6: Evaluation
S7: Alternative
S8: Framing
S9: Grouping
S10: Unity
S11: Objectivity
S12: Tentativeness
S13: Metalanguage
SP: "Space" Data Holistic Score

Table B-5

Intercorrelation of the Analytic Subcriteria Scores and the Holistic Scores for the "Farm" Data

	Variables													
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	FR
F1	1													
F2	.82	1												
F3	.75	.85	1											
F4	.74	.83	.88	1										
F5	.72	.84	.89	.89	1									
F6	.73	.81	.81	.86	.85	1								
F7	.61	.73	.67	.69	.66	.79	1							
F8	.82	.85	.79	.78	.79	.76	.67	1						
F9	.82	.84	.83	.80	.80	.77	.64	.87	1					
F10	.79	.87	.86	.81	.88	.79	.68	.87	.88	1				
F11	.78	.68	.75	.75	.74	.68	.52	.77	.77	.76	1			
F12	.72	.77	.80	.77	.78	.72	.66	.80	.79	.80	.75	1		
F13	.56	.69	.68	.67	.66	.63	.62	.67	.64	.69	.53	.70	1	
FR	.69	.73	.68	.68	.66	.65	.58	.75	.72	.73	.67	.71	.69	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 16. The Variable names are as follows:
F1: Adequacy if Information
F2: Richness of Information
F3: Relationships
F4: Inferences
F5: Synthesisn
F6: Evaluation
F7: Alternative
F8: Framing
F9: Grouping
F10: Unity
F11: Objectivity
F12: Tentativeness
F13: Metalanguage
FR: "Farm" Data Holistic Score

Table B-6

Intercorrelation of the Analytic Main Criteria Scores, the Analytic Total Scores, and the Holistic Scores for the "Space" Data

	Variables				
	SA	SB	SC	ST	SP
Content/Thinking Scores (SA)	1				
Organization Scores (SB)	.93	1			
Style/Tone Score (SC)	.90	.91	1		
Analytic Total Scores (ST)	.97	.98	.97	1	
Holistic Scores (SP)	.82	.80	.76	.82	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 16. The analytic main criteria scores were based on the mean scores of the two rater groups' ratings.

Table B-7

Intercorrelation of the Analytic Main Criteria Scores, the Analytic Total Scores, and the Holistic Scores for the "Farm" Data

	Variables				
	FA	FB	FC	FT	FR
Content/Thinking Scores (FA)	1				
Organization Scores (FB)	.92	1			
Style/Tone Score (FC)	.88	.89	1		
Analytic Total Scores (FT)	.97	.97	.95	1	
Holistic Scores (FR)	.74	.77	.78	.79	1

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 16.

Table B-8

Intercorrelation of the Analytic Main Criteria Scores for the "Space" Data and the Analytic Main Criteria Scores for the "Farm" Data

	Variables					
	SA	SB	SB	FA	FB	FC
SA	1					
SB	.940	1				
SC	.913	.910	1			
FA	.723	.737	.727	1		
FB	.747	.737	.750	.922	1	
FC	.779	.782	.761	.882	.885	1

Note. The above intercorrelation matrix (n = 96) was used in the confirmatory factor analysis as shown in the Table 17. The analytic main criteria scores were based on the mean scores of the four raters' ratings. The Variable names are as follows:
SA: Content/Thinking Scores for the "Space" data
SB: Organization Scores for the "Space" data
SC: Style/Tone Scores for the "Space" data
FA: Content/Thinking Scores for the "Farm" data
FB: Organization Scores for the "Farm" data
FC: Style/Tone Scores for the "Farm" data

Table B-9

Intercorrelation of the Analytic Main Criteria Scores and the GRE Verbal Scores with the Syntactic Characteristic Variables for the "Space" Data

	Variables			
	SA	SB	SC	VL
TW	.76	.77	.79	.73
TU	.44	.42	.44	.35
WO	.28	.30	.35	.31
WI	.38	.36	.37	.40
WC	.54	.53	.57	.51
WF	.60	.60	.65	.60
MT	.44	.45	.43	.48
MO	.07	.10	.13	.15
MI	.35	.32	.32	.35
MC	.36	.35	.37	.35
MF	.38	.38	.41	.41

Note. The above intercorrelation matrix (n = 96) was used in the multiple regression analysis as shown in the Table 19. For the Variable names, see Table B-1 and Table B-8.

Table B-10

Intercorrelation of the Analytic Main Criteria Scores and the GRE Verbal Scores with the Syntactic Characteristic Variables for the "Farm" Data

	Variables			
	FA	FB	FC	VL
TW	.79	.79	.67	.57
TU	.55	.52	.33	.22
WO	.22	.17	.17	.19
WI	.32	.39	.40	.39
WC	.42	.50	.43	.48
WF	.48	.53	.48	.52
MT	.28	.35	.45	.42
MO	-.17	-.15	-.04	-.03
MI	.25	.31	.34	.31
MC	.17	.26	.27	.37
MF	.09	.17	.25	.31

Note. The above intercorrelation matrix (n=96) was used in the multiple regression analysis as shown in the Table 19. For the Variable names, see Table B-1 and Table B-8.

Table B-11

Intercorrelation of the Analytic Total Scores and the GRE Verbal subsection scores

	Variables						
	ST	SP	FT	FR	SN	DV	RC
ST	1						
SP	.816	1					
FT	.798	.735	1				
FR	.826	.860	.791	1			
SN	.717	.789	.697	.755	1		
DV	.719	.750	.653	.794	.862	1	
RC	.762	.807	.749	.830	.829	.800	1

Note. The above intercorrelation matrix (n=96) was used in the confirmatory factor analysis as shown in the Table 21. The Variable names are as follows:
ST: Analytic total scores for the "Space" data
SP: Holistic scores for the "Space" data
FT: Analytic total scores for the "Farm" data
FR: Holistic scores for the "Farm" data
SN: GRE Verbal Sentence Completion subsection scores
DV: GRE Verbal Discrete Verbal subsection scores
RC: GRE Verbal Reading Comprehension subsection scores

Table B-12

Intercorrelation of the Analytic Total Scores and the GRE Verbal Subsection Scores, and GRE Reasoning Subsection Scores

	Variables								
	ST	SP	FT	FR	SN	DV	RC	AR	LR
ST	1								
SP	.826	1							
FT	.776	.732	1						
FR	.807	.858	.801	1					
SN	.704	.748	.661	.726	1				
DV	.676	.730	.628	.719	.808	1			
RC	.734	.765	.696	.764	.801	.774	1		
AR	.311	.315	.341	.366	.385	.384	.416	1	
LR	.628	.679	.624	.683	.752	.680	.769	.474	1

Note. The above intercorrelation matrix was used in the confirmatory factor analysis as shown in the Table 21. The above intercorrelation matrix (n = 205) was based on the original ETS sample (n = 205). The Variable names are as follows:
AR: GRE Analytical Reasoning subsection scores
LR: GRE Logical Reasoning subsection scores
(For the other variables' names, see Table B-11)

Table B-13

Results of Factor Analysis using 13 Analytic Subcriteria Scores

(a) Rotated Factor Pattern (PROMAX)

(standard regression coefficients)

Variable	Factor 1	Factor 2	Factor 3
Adequacy of information	.392	.334	.302
Richness of information	.328	.428	.301
Relationships	.378	.234	.447
Inferences	.604	.132	.299
Synthesis	.582	.204	.250
Evaluation	.651	.245	.135
Alternatives	.550	.346	.090
Framing	.197	.637	.203
Grouping	.398	.507	.131
Unity	.304	.480	.274
Objectivity	.188	.208	.585
Tentativeness	.235	.339	.436
Meta language	.175	.552	.252

(b) Variance explained by each factors eliminating other factors

Factor 1	Factor 2	Factor 3
.824	.737	.513

(c) Eigenvalues of the reduced correlational matrix.

	Factor 1	Factor 2	Factor 3
cumulative %	98.43	99.55	100.40

Note. The analytic subcriteria scores were based on the total scores of the two Topics. For the above factor analysis, the original ETS sample (n = 212) was used.

Table B-14

Relationships between three analytic main criteria:

Maximum likelihood estimates

(a) Factor Loadings

Variables/ Factor	Content/Thinking	Organization	Style/Tone
Adequacy of information	.937	0	0
Richness of information	.957	0	0
Relationships	.958	0	0
Inferences	.948	0	0
Synthesis	.947	0	0
Evaluation	.947	0	0
Alternatives	.903	0	0
Framing	0	.951	0
Grouping	0	.955	0
Unity	0	.970	0
Objectivity	0	0	.890
Tentativeness	0	0	.924
Meta language	0	0	.897

(b) Factor Intercorrelations

	Content/Thinking	Organization	Style/Tone
Content/Thinking	1.00		
Organization	.978	1.00	
Style/Tone	.977	.975	1.00

APPENDIX C

WRITING TASKS

1. "Space" Topic.

TIME - 30 MINUTES

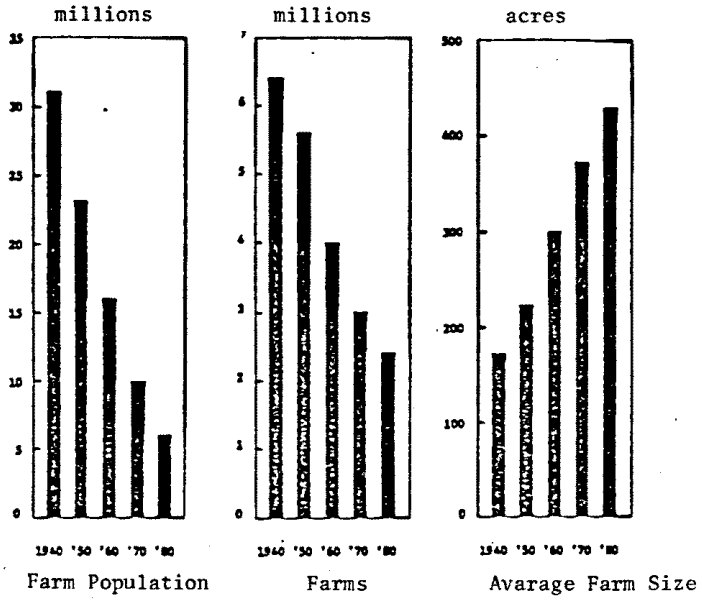
Some people say that exploration of outer space has many advantages; other people feel that it is a waste of money and other resources. Write a brief essay in which you discuss each of these positions. Give one or two advantages and disadvantages of space exploration, and explain which position you support.

THIS SPACE MAY BE USED FOR NOTES.

2. "Farm" Topic.

TIME - 30 MINUTES

CHANGES IN FARMING IN THE U.S.: 1940 - 1980



Suppose that you are writing a report in which you must interpret the three graphs shown above. Write the section of that report in which you discuss how the graphs are related to each other and explain the conclusions you have reached from the information in the graphs. Be sure the graphs support your conclusions.

THIS SPACE MAY BE USED FOR NOTES.

VITA

Young Mok Park was born in Chinju city, Korea on 27 January 1948. He was graduated from Chinju High School in 1966. He entered Seoul National University in 1966 and received his Bachelor of Arts degree in 1970. After completing three years of military service, he was a high school teacher of the Korean language. He was accepted into graduate school at Seoul National University in 1976, completing his Master's degree in Korean language education in 1978.

After working as a specialist at the National Council of Korean Language for one year, he worked as an educational researcher at the Office of Supervision, Ministry of Education in Korea until 1983. During this time he participated in establishing several educational plans for elementary and secondary schools such as the national plan for normalization of school education, the credential system in high schools for college-entrance examination, and national programs to improve classroom teaching in the elementary and secondary schools.

He received the Ph.D from the University of Illinois at Urbana-Champaign in January of 1987. At the University of Illinois, he worked as a research assistant at the Curriculum Laboratory until 1986. His current research interests include reading comprehension process, writing process, measurement of writing ability, and classroom teaching methods in the ill-structured area of the language arts.