AN ANALYSIS OF THE PERFORMANCE OF NON-NATIVE UNIVERSITY STUDENTS ON AN ENGLISH PROFICIENCY EXAMINATION

By

JEANA MARIE SNODDERLY JOHNSON

Bachelor of Arts

Louisiana State University at New Orleans

New Orleans, Louisiana

1971

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF ARTS
May, 1989

AN ANALYSIS OF THE PERFORMANCE OF
NON-NATIVE UNIVERSITY STUDENTS
ON AN ENGLISH PROFICIENCY
EXAMINATION

Thesis Approved:

Pair Shoesey
Thesis Advisor

Carol Lynn Moder

Richard P. Battiger

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

Education is a wonderful thing; it allows one to see where once he was blind. I want to thank my professors and mentors at Oklahoma State University for allowing me to see.

Especially, I would like to thank Dr. Richard Batteiger for giving me the opportunity to gain special insight into this subject and for his patient listening and useful commenting as I tried to clarify my ideas. I also want to thank Dr. Carol Moder for her assistance in compiling and interpreting the statistical data and for continuing to smile encouragement even as I unrolled miles of spreadsheets across her desk. I am very grateful to my advisor, Dr. Ravi Sheorey, who probably is unaware how he has helped me pull order out of chaos.

Finally, I want to express my deepest appreciation to my husband David. His herculean efforts in helping me compile the statistical data was nothing compared to his constant encouragement, understanding, and love during this project. Thanks, Dave. I won't forget it.

TABLE OF CONTENTS

Page

LIST OF TABLES

NOMENCLATURE

Agr Pro	Pronoun agreement
Agr S/V	Subject/Verb agreement
Body	Body, with supporting main points
Cap	Capitalization
Comma	Punctuation, commas only
Concl	Conclusion
Conn	Connective elements
Details	Details supporting the main points
DisSc	Discourse Analysis Score
Err/Wd	Errors per word
EssayNo	Proficiency exam by student numer
F1, F2,...	Failing students by assigned number
ExScore	Holistic composite exam score
Failing	The group of 19 exams that received failing holistic scores
%Fail	See %Pass
Frag	Sentence fragment
IncConst	Incoherent constructions
Intro	Introduction
L1	Native language
L2	Second language
Mean	Mean scores for that column
Mech	Mechanical category

N-e	Normalized error
N/P	Noun/Pronoun category
NoWords	Number of words
Number	Singular, plural of nouns, pronouns
Organ	Organization
P1, P2,...	Passing students by assigned number
Parallel	Parallel constructions
Passing	The group of 20 exams that received passing holistic scores
%Pass	TABLE VII--the percent of the total passing discourse points
	TABLE VIII and TABLE XII--the percent of the total points received by the passing group
Poss	Possession
Punct	Punctuation, excluding commas
rScore	Correlation Coefficient, all correlated with the actual composite exam score
Range	The maximum minus the minimum score
Ratio	For error analysis--the ratio of TotN-e Failing to TotN-e Failing
	For discourse analysis--the ratio of Mean Passing to Mean Failing
RunOn	Run-on sentence
Sent	Sentence category
Spelling	Spelling
StdDev	Standard Deviation
Tense	Verb tense
Th. State	Thesis statement
TotErr	Total number of errors
%TotErr	Percent of total errors

TotN-e	Total number of Normalized errors
UnclearPro	Unclear pronouns
Verb	Verb category
Vocab	Vocabulary category
WC Art	Word choice, articles
WC Aux	Word choice, auxiliaries
WC Conn	Word choice, connective elements
WC Mod	Word choice, modifiers
WC N/V	Word choice, nouns, pronouns, verbs
WC Prep	Word choice, prepositions
W Form	Word form
W Form-ing-ed	Word form, -ing and -ed endings only

CHAPTER I

INTRODUCTION

The Problem

According to Kaplan (1987), humans are predisposed to speak; we are not predisposed to write. He reports that evidence shows the capacity to speak entered human DNA some 100,000 years ago, but that writing was invented only 10,000 years ago. In addition, he says, not only is the capacity to write not built into our genetic code, it is not universally distributed in the species.

The premise that humans are not predisposed to write is an important obsevation when discussing proficiency in writing. In other words, since the capacity to write is not genetic, we must have to learn it. And since writing is not evenly distributed among humans, we must not all write in the same way or with the same facility, if indeed we have learned to write at all.

According to Kaplan (1987), writing has caused dramatic changes in those cultures which have developed it. We have changed our attitudes toward fact and truth. We don't have to hold information in our memories anymore; consequently,

being able to write information has encouraged us to be more accurate. Another important change occurred when the printing press was invented some one thousand years ago: Wide dissemination of information was possible. It has been only within our lifetime that a third important change occurred: We have developed automated word processing. Now we can get huge volumes of information instantaneously.

However, any new change puts constraints on writing even as it strives to break down barriers to communication. Writing styles change, language usage evolves, and cultures continue to redefine what constitutes good writing. How, then, can we successfully determine written proficiency for native speakers of our language, much less for non-native speakers?

Colleges and universities have a reason to examine writing proficiency in second language learners. The problems facing second language learners at the university level, which is the focus of the present study, are varied. Richards (1974b) discusses many characteristics these students have in common. They are international students here in the United States for the first time to complete their college education. They have learned their English as a foreign language (EFL) in their home country where English usually is not a viable second language. They have been taught English in a formal setting, a classroom, perhaps. Their study of the language has had a cultural objective, not a societal or an economic one. For instance, students

might have been taught cultural lessons about the best art and music in an English-speaking country rather than typical discourse schemas of American university students as they might occur in the student union; or they might have been given poetry or classical fiction to read instead of current technological, business, or academic articles reflecting American thought or achievement. Consequently, they have not been highly motivated to learn the language thoroughly in an appropriately communicative manner since they would not have had the need to use it in their own country very often. If they did have an opportunity to use English, they would have used a local form, not an overseas standard. This often has meant they have acquired inaccurate forms, pronunciation, and discourse strategies and inadequate vocabulary.

Speaking a local form of English often promotes a kind of "interlanguage." Selinker (1972) coined the term "interlanguage" to mean a level of language competence on its way toward being, but not yet achieving, native-like competence. It is an honorable level of proficiency since the second language learner is in the process of perfecting second language rules as he uses the language. Nevertheless, he may eventually reach a point where he is satisfied with the level of "interlanguage" he has mastered since he is able to communicate well with it, even though it is less than native-like. His "interlanguage" level in that

case could be anywhere on a continuum of little competency to native-like competency.

Though it appears Selinker intended "interlanguage" to describe spoken language, the same term could be used to describe the various proficiency levels of written language. A second language learner coming to an American university for the first time may have been taught the English writing rules and rhetoric patterns of another English speaking country rather than American writing rules and patterns. In those cases, the learner might have acquired different spelling forms, syntactical constructions, or rhetoric patterns unfamiliar or unacceptable to American university instructors. Learning a "foreign" set of writing rules and patterns isn't the only problem; the second language learner may have incorrectly or incompletely learned the writing rules and patterns of his own native language. These situations might cause the learner to develop a complicated kind of writing interlanguage: One that contains a mixture of native language (L1) and second language (L2) rules and patterns.

The problem for U.S. colleges and universities is that international students arrive in all stages of spoken and written interlanguage, even though they have achieved an acceptable score on the required TOEFL test. By coming to the United States, these EFL students have suddenly become ESL students. That is, they have come from a background where English has been learned as a foreign language to a

situation where they must learn English as a second language. According to Richards (1974b), an ESL student

> ...is subject to all the censure of the society who recognizes a deviant variety of English as opposed to noticing just mistakes in a developing interlanguage. He must conform immediately, as an ESL student would, and immediately erradicate all his problems. It is doubly hard on the university student, who is often not encouraged by his American classmates to socialize with them, or by the university programs which often separate him from native speakers in scheduling English classes. Yet, without socialization with natives, he cannot quickly eradicate his "deviant" variety of English (p.90).

Further English language instruction at the university is usually necessary, but it is often not enough to help them achieve a proficiency level acceptable to American university academic standards. The international student often must pass an English writing proficiency test as a university graduation requirement, and he must meet the same standards of academic proficiency as do native speakers of English.

According to Holderer (1988), at Oklahoma State University during a seven month period, 58 percent of the non-native speakers failed an English writing proficiency examination required of all university students before graduation; during the same period, only 11 percent of the native speakers failed the test. Non-native speakers failed the writing test at a rate of nearly five to one over native speakers.

The Hypothesis

If the international student cannot produce native-like English responses to an essay question, then he must be using some other form of English (Selinker's interlanguage), one that makes sense according to his level of competence in understanding and using the writing rules he has observed. If Richards's (1974b) description of the ESL student is accurate, then our EFL/ESL university students may not be getting all the help they need to achieve an acceptable level of academic proficiency. In addition, if American university standards require these EFL/ESL students to pass written proficiency examinations for graduation with a university writing proficiency level, then it is even more important to correct their "deviant variety of English." One way to do this is to analyze the performance of non-native speakers on a written English proficiency examination and apply what we learn to improve our ESL instructon and testing programs.

This study analyzes the performance of non-native writers on a university writing proficiency examination in order to attempt to identify the features which apparently cause a student to pass or fail.

Among the questions to be explored in this study are the following:

1. Are there features between those who pass and those who fail which might clearly

determine proficiency?

2. What kinds of distinctions between those who pass and those who fail would a sentence error analysis and a discourse analysis show that a holistic exam score would not show?

3. How do the sentence error and discourse analyses scores compare to each other and with holistic scoring as a test of general proficiency?

4. What do proficiency raters seem to attend to when judging proficiency.

CHAPTER II

REVIEW OF LITERATURE

Introduction

Interest in evaluating writing proficiency for both native and non-native students has been growing recently. The results are used for a variety of purposes: School administrators rely upon them for placement decisions or as a graduation requirement; teachers use them to help improve, refine, direct, and shape their students' writing abilities and attitudes; researchers use them in order to continue to build upon the body of knowledge needed to understand how we communicate through writing and how we can best determine proficiency.

According to Raimes (1987), most research has been done with native speakers of English and "the findings and implications of these studies have been generalized to second-language students" (p. 439). Troubling to Raimes is that most research seems to stress the similarities between native and non-native writers, and little has been done to research the differences. Similarily, she notes, little has been done to study the differences between skilled and unskilled writers and, in fact, to explain how we presume to

categorize ESL students as such. While Raimes (1985) observes we are not lacking in second language writing studies, she feels the case studies are done with such a limited number of subjects that it "makes it difficult to form conclusive generalizations" (p. 231).

Raimes's observations influence the present study which is primarily concerned with whether features can be found which seem to lead second language university students to pass or fail writing proficiency tests. Secondarily, it will explore those measurable differences shown by a two-level detailed analysis when compared to a one-level holistic scoring, how the exams' analytical error/discourse analyses scores compare with each other and with their holistic scores as a test of general writing proficiency, and whether raters seem to be paying attention to isolated features during scoring sessions.

A review of literature was done to determine the reliability of holistic and analytical scoring when used to evaluate writing proficiency, and whether or not any drawbacks they contained could be compensated for in the present study. Furthermore, research dealing with error and discourse analysis was studied to determine the kinds of features important to a two-level analysis such as the one done here.

Few current empirical studies were uncovered in the review of literature which directly relate to the kind of scoring or analyses done in the present study. Therefore,

also presented in this chapter is an overview of that literature which seems to be making a valuable contribution to the understanding of proficiency scoring and error/discourse analysis in second language learning.

Research on Global Scoring Methods

Methods of writing assessment have been and still are being hotly debated among educators and researchers. While the search for reliable, more objective scoring methods continues, most writing proficiency today is being determined by the use of three global scoring methods: holistic, analytical, and primary trait. Stiggins and Bridgeford (1983) use the following description of these three scoring methods:

> Holistic scoring calls for the reader to rate overall writing proficiency on a single rating scale. Analytical scoring breaks performance down into component parts (e.g., organization, wording, ideas) for rating on multiple scales. And primary trait scoring requires rating of attributes of performance unique to a particular audience and writing purpose (e.g., persuasiveness, awareness of audience) (p. 26).

Two of these methods, holistic and analytical, are central to the present study. These two kinds of scoring give a fast, overall impression of the writing sample and are usually reliable given trained raters, time, and a clear scoring guide; but they also have some drawbacks.

Of all writing assessment methods in use today, holistic scoring, according to Perkins (1983), "has the

highest construct validity when overall attained writing proficiency is the construct to be assessed" (p.652).

Yet, holistic scoring has had much criticism in the past. Burt (1917) noted that "no other form of examination leads to such inconsistent marking" (p. 55).

Valentine (1932) remarked that "...extraordinary variations occur between the marks of different examiners" (p. 26).

Looking for the most reliable global scoring method, Cast (1939) compared four different methods of marking compositions written by forty British school girls. She tested the "general impression " (holistic), the analytic, the achievement, and the individual methods, finding the analytic method slightly superior.

Analyzing different kinds of rater's scores, Diederich, French, and Carlton (1961) found wide diversity in judgement among sixty professionals from varying fields when they subjectively graded 300 papers written by college freshmen from three different schools.

Even today, threats to reliability continue to plague holistic scoring. According to Perkins (1983), the rater's overall impression may be based on experience with scoring other student's writing or on an absolute standard observed in professional writing.

> Such an evaluation can, therefore, be highly subjective due to bias, fatigue, internal lack of consistency, previous knowledge of the student, and/or shifting standards from one paper to the next (Perkins 1983, p. 653).

Consequently, over the years, holistic scoring has undergone intensive research designed to reduce the threats to reliability its critics claim it has. Lado (1961) found that holistic scoring can be effective, but requires highly trained personnel and time.

Testing Lado's observations, Braddock, Lloyd-Jones, and Schoer (1963) reported interrater reliability coefficients as high as .90 using holistic scoring.

In addition to needing trained raters and time, Harris (1969) found that when being scored holistically, writing samples must be read by several experienced readers using a scoring guide of the general components to be tested. Those components, Harris says, should include content, form, grammar, style, and mechanics.

Comparing holistic scores to Hunt's (1965) T-unit concept (a count of T-units, or each main clause, as an indicator of writing maturity), Kaczmarek (1980) found holistic scoring by teachers not only yielded substantial reliabiltiy estimates, but strongly correlated with objective scores.

Current thought is demonstrated by White's (1982; 1984) work. White's (1982) eight-year study of the effectiveness of features of writing programs found, after examining the holistic scoring procedure of thousands of California's placement tests for college and university in-coming freshmen, very high rater reliability estimates (nearly .90). White (1984) developed a substantially reliable

holistic scoring method used currently by many American universities and colleges. He suggests writing programs directors should define the kinds of skills different kinds of writing prompts ask for, then carefully compose the writing prompts to tap those skills, reduce the variability in the scoring process by using a scoring guide based upon the skills needing to be assessed, and use only holistic scoring with trained raters working under special conditions.

Yet, even White (1984) admits to some limitations to holistic scoring: It has no meaningful diagnostic information beyond comparative ranking, scores cannot be easily normalized since they are always relative to writing prompts and to the student body, and it yields only approximate levels of proficiency since if scored again by the same readers, the same examinations would probably receive different scores. He suggests that an absolute holistic proficiency cut-off score might not even be realistic; a band of scores might be more appropriate for the pass/fail line.

Cooper sums up holistic evaluation:

> Where there is commitment and time to do the work required to achieve reliability of judgment, holistic evaluation of writing remains the most valid and direct means of rank-ordering students by writing ability. Spending no more than two minutes on each paper, raters, guided by...holistic scoring guides..., can achieve a scoring reliability as high as .90 for individual writers. The scores provide a reliable rank-ordering of writers, an ordering which can then be used to make decisions about placement, special instruction, graduation, or grading (cited in Perkins 1983, p. 655).

While holistic scoring is more often used as "a tool for certification, placement, proficiency, and research testing" (Perkins 1983, p. 653), analytical scoring is primarily used for "correlational research, exemption, growth measurement, prediction, placement and program evaluation uses" (Perkins 1983, p. 656).

Bloom, Hastings and Madaus (1971) report analytical scoring helps the rater to focus on the features to be evaluated and therefore helps "to determine the degree of mastery of a given learning task and to pinpoint the part of the task not mastered..." (p. 61).

Analytical scoring is also recommended by Heaton (1975) for situations in which a single rater is used. He notes that this method attempts to separate the various features of a compostion for scoring purposes. These features are scored individually, and a composite score is reached. For the classroom teacher, who often cannot call upon multiple raters, an analytical scoring scheme can help distinguish individual features of writing which might need further instruction.

Cooper(1977) observed, "Where a criterion measure is required in a research study, raters can use an analytical scale to score each student's writing" (p. 17).

Agreeing with Kaczmarek (1980), Zughoul and Kambal (1983) found that analytical scoring is better than holistic scoring in certain situations. After developing a detailed analytic testing procedure for EFL compositions, they

compared it with a holistic scoring of 90 writing samples. Their conclusion was that the analytic method was able to better predict which features, when mastered, were indicative of a basic, intermediate, or advanced writer.

However, as with holistic scoring, analytical scoring can have serious disadvantages. Cooper (1977) reports that an analytical scale is time-consuming to prepare. He describes six steps in this procedure: First, features to be analyzed must be derived from a large corpus of published and original student writing; second, the scale must be field-tested, after which the features may be modified; third, the high, mid, and low quality levels for each feature must be described; fourth, points must be anchored on a scoring line; fifth, raters must practice using the scale; and finally, intra- and interrater reliability must be measured. Cooper (1977) notes, with analytic scoring, that the scoring weight of a particular category is "not sensitive to the variations in purpose, speaker role, and conception of audience which can occur in pieces written in the same mode" (P.14).

Furthermore, White (1984) criticizes the validity of analytic scoring as a method which breaks into parts and examines "art" (writing) which should be taken as a whole. He reports that raters cannot agree on sub-skills to be assessed and that the analytical scoring process is a very complicated process.

Perkins (1983) comments that, as with holistic scoring, some raters bias the standard by trying to adhere to an absolute code of published professional writing and, therefore, discount the standards of the student writing corpus. Secondly, he continues, features to be analyzed are isolated from the context and scored separately, while the whole of the text is ignored. He sees this as a challenge to the validity of the scores. Perkins's third criticism of analytical scoring is that the categories themselves are often vague and certainly arbitrary. He feels different raters value different aspects of writing. However, Perkins concludes , "With enough time and commitment on the part of the graders, reliable scores can be obtained" (p.658).

Research on Error/Discourse Analysis

The literature on error analysis seems to indicate that this is a broad term primarily intended to mean the study of error in language whether it be at the word, sentence or overall discourse level. It does not embrace a particular kind of method, though often it entails quantitative, rather than qualitative, techniques. A counting of the sentence surface errors, such as is used in the present study, is one such quantitative technique. However, an analytical analysis of discourse features, a somewhat qualitative technique used in this study, could also be used.

Early reseach includes Corder's (1967) proposal that L2 errors are not just interference from L1, but evidence of the system the learner is using; he has not acquired the right one yet. Corder distinguishes between mistakes--slips of tongue due to such things as fatigue, memory limitations, or psychological factors--and errors, which occur systematically and consistently throughout the learner's speech or writing. Physical states, mental states, and memory lapses cause learners to make mistakes, but do not cause errors. Corder believes a search for the sytematic error, not the occasional mistake, would give a better indication of competence. He further believes that a systematic study of a learner's errors will result in discovering the system that he is currently using and will help educators plan the appropriate instruction needed to move the learner's system closer toward the target system.

Richards (1971) cites Coulter (1968) when he reports that errors involve strategies of learning. The second language student wants to make learning the new language easier on himself; therefore, he will learn only that which he needs to communicate with the native speaker. His second language can fossilize here. The learner is satisfied with the amount of L2 he has learned, even though it is plagued with errors. Often, Richards says, these errors cannot be corrected by instruction.

Selinker (1972) calls Corder's and Coulter's deviant system "interlanguage." He believes interlanguage is a good

sign. It implies that the learner is formulating hypotheses about the rules of L2 and is testing them to see if they work. Testing hypotheses is how he refines his system. However, mistakes occur when the learner is in a state of anxiety or lacks concentration. Selinker agrees with Corder that a systematic study of the learner's errors will help him advance his knowledge of the rules. However, he says that perhaps a mere 5 percent of adult second language learners ever achieve native competence.

Corder (1974) believes that interlanguage has a grammar; that interlanguage shares rules with L1 and L2, but is a different language from either of those two--an idiolect; that interlanguage is interlanguage precisely because its rules are not known yet. The second language learner's interlanguage is criticized as deviant because it fails to follow a known rule. Yet, it is following rules--hypothesized rules, unknown to the criticizer. Corder feels that an error analysis of a learner's interlanguage, both good and bad sentences, will help the instructor discover the rules. His technique is to compare a sample of the learner's interlanguage with a translation of it in the learner's L1 and then to explain the differences.

Richards (1974b) suggests four reasons why errors in interlanguage occur. First, the second language learner overgeneralizes rules he has learned from other structures in L2. For instance, he knows that except for the third

person, all present tense English verb forms have a zero morpheme. Therefore, by leaving out the third person -s, he does not have to go to the trouble of making the verb agree with the subject. Second, the second language learner is ignorant of rule restrictions. He misuses a previously learned rule in a new situation or applies rules learned by rote, but not clearly understood. For instance, "He showed me the book" becomes "He explained me the book." In this example, he is unaware that some verbs do not take indirect objects. Third, the learner incompletely applies the rules. Richards says the second language learner is not concerned so much with accuracy as with communication. Finally, Richards suggests that the learner hypothesizes false concepts inadvertently caused by teaching methods or materials. For instance, the teacher might present material demonstrating the rules he/she is teaching, but neglects to present some general exceptions.

Jain (1974) agrees with Richard's observations that generalization creates interlanguage. He also believes that the learner's wanting to reduce speech to a simpler system is a universal learning strategy. In addition, he feels that rules, when taught, should be expanded to avoid errors in special contexts. Jain acknowledges that sub-categorization is difficult; but to a second language learner, it does not seem important that rules don't apply in occasional contexts. If he does not find an error significant, he will continue to generalize. According to Jain, some rules will

always be hypotheses because the target language has too many cues, too many exceptions to the rule, for the learner to formulate a complete rule. Jain uses the example of missing English articles. The learner will formulate a partial rule for the use of articles that results in overall good communication; thereafter, he stops trying to formulate a complete rule once he is satisfied with his level of interlanguage.

Shaughnessy (1977) has given the most convincing demonstration of the difficulty that second language learners have with errors. Though she was studying native speakers of English dialects who were enrolled in her basic writing classes, she has allowed ESL researchers and teachers to see that they must look for patterns of error which often reveal the faulty logic of the erroneous rules and strategies which the second language learner is applying. Shaughnessy believes those patterns of error can only be revealed by analyzing both what the writer does right and what he does wrong. For instance, she cautions that the basic student is not as concerned with style as with mechanical correctness; yet, syntax is largely concerned with style. If a student has trouble with syntax, he has trouble with the relationship between words in a sentence. Therefore, his sentences seem incoherent to the reader.

Cummins (1980) proposes the theory that language proficiency is separated into interpersonal and

cognitive/academic components he calls BICS (Basic Interpersonal Communicative Skills) for daily language needs and CALP (Cognitive/Academic Language Proficiency) for academic language settings. According to his hypothesis, BICS is learned on the playground and CALP, in the classroom. If a second language learner makes errors in an academic situation, it may be because his CALP has not been sufficiently developed. Insufficient CALP development in L1 can manifest itself in L2. One can infer from Cummins theory that CALP, where it concerns writing, must develop in the classroom. It isn't learned on the playground.

Before Cummins's theory on two-fold language proficiency, another theorist proposed a related problem that could be inherited from the L1. Twenty-three years ago Robert Kaplan (1966) stunned linguists by proposing that rhetorical thought patterns are determined by one's culture. He suggested particular graphic forms as representative of different rhetorical structures of the world's languages. Since then, Kaplan (1987) has admitted that his earlier view was too strong. He explains that while critics agree that there are important differences in the way languages identify discourse topics and the way topics are developed, they are correct in disagreeing with the nature of the rhetorical forms. Kaplan insists, however, that different languages have different rhetorical preferences, and do not choose all forms with equal frequency, nor are they interchangeable without putting constraints upon the text.

Furthermore, in learning a second language, the learner does not possess the large inventory of vocabulary and syntax possibilities that the native speaker does. For instance, the learner does not recognize constraints on language or sociolinguistic implications of certain vocabulary words or idioms.

Kaplan suggests that the reason we do not have many empirical studies on written text error and discourse analysis on enough samples to make the conclusions meaningful is that this type of research happens to take an inordinate amount of time and that the work is extremely tedious. Nevertheless, he proposes that writing samples need to contain more than 400 words to yield meaningful conclusions.

Finally, Kaplan concludes, "No soundly based theoretical model for the study of written text exists" (p.19) He feels a definition of that model should include semantics, grammar, rhetoric, and audience. No single level of analysis will be adequate in describing written text, Kaplan reports. A truly analytical study of written text must deal with several areas at once.

Bhatia's (1974) study of ten second language students at the University of Dehli is one of the few empirical studies uncovered in this review of literature which is directly related to the present study. She believes that an analysis of errors that have actually occurred gives better and more reliable results upon which to determine

proficiency and appropriate instruction than than would a hypothesized or generalized schema of probable errors which might entail reteaching all language features, whether or not they have been mastered. Her subjects were in an intermediate stage of interlanguage and were asked to respond to a student-chosen topic which required narration or description. No special vocabulary requirements hindered the student's essentially free expression. No test anxiety preconditioned the writing assignment since it was offered as part of their regular classwork. They were, however, under a 55-minute time limit, the usual time duration of a class hour.

Their writing samples averaged 250 words. They were analyzed using two broad catgegories: errors of a grammatical nature within the sentence, which Bahtia called Mechanics; and errors of paragraph development, which she called Organization. Each category was subdivided into individual features which were then described. For instance, Mechanics included such features as failure to give a verb to each clause, subject-verb agreement, articles, prepositions, and plurals; features listed under Organization included relevance, order, clarity of construction, adequate development, and originality.

Bhatia's study revealed a 40 percent error frequency in verb forms and tense sequences and a like percentage of error in article use. In addition, she found that students were 100 percent deficient in originality and adequate

development as determined by her feature definitions; and they were 70 percent deficient in clarity of construction. Bhatia's conclusion included the following:

> an error-based analysis gives reliable results upon which remedial materials can be constructed; a study of the percentage values of different errors gives us an insight into the relative significance of a given error; [and] a course based on the frequency of errors will enable the teacher to teach at the point of error and to emphasize more those areas where the error freguency is higher... (p. 349).

A second empirical study was undertaken by Rollins (1985) on the analysis of the writing of thirty native but limited English-proficient students. These subjects were all bilingual, having learned their native English as a dialect or as a language different from the one used in the home. Most of the subjects were English/Spanish speakers. Rollins analyzed sixty compositions in which each subject wrote responses to two different tasks. Task I was to write about a personal quality the writer would like for others to know about. Task II was to write a comparison/contrast response about the pictures of two advertisements. Rollins was interested in the amount of Spanish language interference in the English writing of these subjects, the kinds of surface errors occurring in sentences, the kinds of discourse errors made in the text overall, and how the subjects' errors compared to their reading levels.

Rollins divided her analysis into Errors and Discourse, thereby completing a two-level analysis of each sample. First, she determined the number and categories of features to be used in analyzing the surface errors in the sentences

of the sixty samples. For example, included in the fifteen features she chose to analyze were errors in sentence fragments, awkward sentences, misspelled words, subject/verb agreement, and word choice. She divided the fifteen features into five categories. She then counted the errors in the use of these features after marking them in the samples. Finally she determined the frequency count of errors by category and task.

For the second part of her analysis, she determined which features would best determine the aim of discourse in the text of the samples. She analyzed the text in both a quantitative and qualitative way. Quantitatively, she counted the total number of words, sentences, and T-units (see Hunt 1965) in each sample. Qualitatively, she evaluated each sample's control of paragraphing, cohesion, and coherence by preparing a descriptive scoring guide of each feature.

Finally, she administered the Nelson-Denny Reading Test, Form E, to her subjects to determine mean grade equivalents.

After analyzing sentence errors, she found misspelled words to be the most frequent error, followed by at least fifty recorded errors each for run-on sentences, punctuation errors, and word choice errors. In addition, she found that among her English/Spanish subjects only a small number of errors overall could be attributed to language interference.

Rollins's discourse analysis showed that the mean length of Task I was 207 words; the mean length of Task II was 195 words. Her subjects wrote 30 percent more sentences in Task I compared to Task II. Consequently, she concluded that the number of sentences does appear to be affected by the topic of the composition. Task I showed more T-units, a concept which she feels shows maturity in writing. Rollins feels the higher number of words, sentences, and T-units in Task I may be directly related to social factors (see Cummins 1980). She suggests that a bilingual student's linguistic maturity is affected by the amount and kind of his exposure to English linguistic schemas. It is through personal experience with these schemas that he is able to build the vocabulary and syntax necessary to speak or write about them in the target language, much like Cummins's concept of BICS and CALP. Task I asked him to write about a personal attribute, a linguistic schema he was familiar with (Cummins's BICS). However, Task II may have asked him to write about a linguistic situation he had no experience with (Cummins's CALP). Therefore, he struggled more with the rhetoric. Regardless of the task, however, Rollins found that her subjects had trouble with paragraphing, cohesion and coherence. She feels this trouble is due to inexperience in writing, indicating that the subject is lacking in those skills.

Finally, Rollins found that reading proficiency seems to be directly related to writing proficiency: the better

the reader, the better the writer. Most of her limited university students scored at an eighth grade reading level, though there were wide variances in the scores. She concludes that reading/writing are complementary skills, just as speaking/listening are. What affects one might affect the other.

Summary

The broad scope of this review of literature reveals current thought on global writing proficiency scoring methods and error/discourse analysis, particularly as they apply to second language learners. The literature seems to support that holistic and analytical scoring techniques both yield reliable results in the measurement of writing proficiency depending upon the reason for testing and the results needed. Holistic scoring requires several experienced and trained raters, a sufficient time allowance to counteract rater fatigue and bias, writing prompts carefully written to tap the skills to be assessed, and scoring guides which adequately reflect the standards to be evaluated. It is best used as a method to quickly test for overall proficiency on large numbers of writing samples. Analytic scoring is best used to diagnose particular or individual problems with writing, and it is easier to use with a single rater who sets his own categories of features to be analyzed.

Critics claim that these scoring techniques, while they may be the most reliable we have, are not foolproof. In fact, Raimes (1987) feels, "There is at present no consensus on valid criteria for measuring skill in writing and thus no clear agreement on the meaning of 'unskilled'" (p. 231). She calls for more research. Less global evaluation methods such as error analysis, on the other hand, have proved to be also less controversial. The literature supports studies incorporating quantitative measures, perhaps in an effort to have the data appear more empirical, less impressionistic. Certainly errors, when isolated by analysis, can be studied for their significance. Yet, even with an error analysis of discourse experts hedge that rhetorical features are more easily discovered by qualitative measures than by purely quantitative ones. Whatever the outcome of the debate on writing proficiency scoring methods, continued research using various combinations of techniques seems to be encouraged. Efforts at research on writing are hampered by the tedious, time-consuming work needed for the analysis of writing samples longer than 400 words.

This survey of literature, while not discovering many empirical studies directly relating to the analysis of university proficiency examinations and non-native students, has revealed a profile of the causes of errors in the second language learner's writing. In proficiency examinations such as those at most universities, his level of competence is often not actually measured, rather his level of

performance compared to American university writing standards is. His performance is affected by time factors, anxiety, memory lapses, social background, and culture. Along with 95 percent of second language learners, he has achieved a certain level of competence, called interlanguage, which yet contains varying degrees of faulty logic about the rules of the second language. He follows universal second language strategies: he generalizes about structures he already knows for use in new situations; he tries to reduce language to a simpler system; he hypothesizes rules and tries them out to refine them. He is mostly concerned with mechanical correctness rather than syntax; if he is concerned with syntax, he is mostly concerned with meaning rather than accuracy of structure. He comes from a culture where thought patterns and rhetorical patterns may differ from those of English, and he may not have become proficient in using them even in L1.

Finally, the literature shows that a systematic study of errors and faulty logic in a second language learner's writing is necessary before his interlanguage level can move closer toward native-like proficiency.

CHAPTER III

THE STUDY

Purpose

English writing proficiency examinations are often required for graduation from American universities. Both native and non-native speakers take the same test. As I indicated in Chapter I, Holderer (1988) found that at Oklahoma State University during a seven month period beginning October, 1987, through April, 1988, 58 percent of the 69 non-native speakers anticipating graduation failed such an English writing proficiency examination; during the same period, only 11 percent of 538 native speakers failed the test.

I decided to analyze a sample of the examinations of the non-native speakers who took that written proficiency test during an eleven month period in 1987 and 1988, which included the seven month period of Holderer's study. I intended to study the writing of both those who failed and those who passed to determine whether these students passed or failed for identifiable reasons.

The proficiency test consists of two essays written in a proctored situation within a two-and-a-half hour time limit. The first essay allows a student to demonstrate his ability to argue persuasively for or against a position within ninety minutes; the second evaluates the student's ability to switch between two kinds of writing on one topic--such as between describing an object and explaining its use--in thirty minutes. The examinations are evaluated holistically by at least two raters who have had their rating ability calibrated independently for each set of writing tasks. Each of the two writing tasks on the examination has its own scoring guide. The task requiring persuasion (Task I) has a six-point rating scale, and the task requiring description/explanation (Task II) has a four-point rating scale. Two raters must agree within one rating point on the score for each task on a particular writing sample. Those two raters' scores are then added for the final task score, and the two final task scores are added for one composite examination score.

A review of literature indicates that to yield conclusive results, studies must analyze a large group of writing samples each containing at least 400 words, and writing samples should be analyzed on at least two levels using a combination of scoring techniques. The writing samples from the university's writing proficiency examination fit the first requirement: all writing samples contain over 400 words, and the test population is large

enough to draw a sample sufficient to yield meaningful conclusions. As an answer to the second requirement, I designed a study which would analyze errors at the sentence level and errors at the discourse level. Each level requires a different type of scoring technique. My count of sentence errors uses a quantitative technique, while my analysis of discourse features is both quantitative and qualitative. This kind of two-level quantitative/qualitative analysis allows me to study distinctions in the data that a holistic score does not reveal. Finally, I compare my analyses scores to the original holistic scores on the writing samples to determine whether they correlate.

Sample

From the exam files available to me in the Freshman Composition Program, I took out all 99 proficiency exams written by international students--identified by a special university-assigned student number--from May, 1987, through April, 1988, for use as my sample. I then numbered each student exam consecutively. Next, I used a random number table to choose twenty examinations that received a passing score and twenty that received a failing score, for a total of 40 examinations. I checked the sample to insure that no duplicate names were included, since students may take the test an unlimited number of times to pass. Finally, I renumbered the students' tests in the sample and listed them

by number instead of by name. In this way, when I analyzed the essays, I would not be influenced by name, nationality, culture, or gender. My final sample was 39 exams with a total of 78 writing samples. (One examination was missing its second essay and was eliminated from the study).

Features Analyzed

Sentence

A reading of the literature indicated that no two experts agree on what features and what categories should be used in any particular analysis of errors at the sentence level. They do agree that an analysis of sentence errors is the easiest kind of analysis to do because those errors are so easy to spot; they also imply that most raters consciously or unconsciously score writing based mainly upon the frequency and kind of errors they see in the sentence (Applebee 1981, cited in Robb, Ross and Shortreed 1986; Zamel 1985).

Two studies influenced my choice of taxonomy. Shaughnessy (1977), in her study of the writing of open admission native speakers of English, indicates that errors are deeply rooted into the sentence because of childhood language systems. In this kind of basic student, not simply the frequency but the type of errors alarm college teachers who are used to the writing systems of freshmen familiar

with a more standard dialect. Shaughnessy suggests beginning with a taxonomy based upon word classes with subdivisions: for instance, verbs, subdivided into verb endings and tense; nouns, subdivided into plural and possessive forms; or pronouns, subdivided into agreement and case. A taxonomy of this kind would uncover kinds of errors second language learners make which might not correspond to those expected by raters who are native speakers of English.

Another study influencing my error taxonomy was done on error gravity. Vann, Meyer, and Lorenz (1984) examined 319 faculty responses to the written sentence errors of non-native speakers of English. They found that most respondents did not judge all errors as equally irritating; and, in fact, they suggested the respondents might have an intuitive hierarchy of errors in mind when they evaluate writing. Vann, Meyer, and Lorenz listed the following errors as least acceptable: word order, it-deletion, tense, relative clause errors, and word choice--all global and/or relatively rare violations for native speakers and, therefore, more likely to interfere with communication (see Burt 1975). However, they stopped short of suggesting that there is a direct relationship between reader comprehension and degree of acceptability.

I decided to design my sentence error taxonomy around Shaughnessy's subdivided word class categories and Van, Meyer, and Lorenz's list of least acceptable errors. The writing samples in this study were analyzed by using five

broad error categories: sentence, vocabulary, mechanics, verbs, and noun/pronoun features. Each category was then subdivided into its particular features. (All examples in this chapter have been taken directly from the proficiency tests used in this study.)

Sentence Category

Included in the Sentence category are incoherent constructions and faulty parallelism. These errors are "global" (Burt 1975) and usually affect the entire sentence. Often it means the student has put words together in a wrong order so that the reader has difficulty or is unable to interpret the meaning of the sentence. For instance, note the following two examples:

> But also the man has been realizing several journeys in the space to do works which help the human been.
>
> After being exposed to the american culture, it really scares me, the idea of being treated unfair and feel the frustration of having my hand tied upon the reality.

Vocabulary Category

In this study, Vocabulary is the category with the most features. However, it only concerns one error--word choice. Vann, Meyer, and Lorenz (1984) named word choice as one of the least acceptable errors; actually, many factors make up word choice errors. Faulty word choice in my Vocabulary

category indicates error in choosing the correct vocabulary word, errors of omission of a word where a native speaker would include it, and inclusion of a word where a native speaker would exclude it. To be listed as an error, the choice must interfere with the reader's comprehension, or the reader must be left with a difficult interpretation of the word. The Vocabulary category separates word choice into eight features: word choice errors in prepositions, connectors, articles, modifiers, auxiliaries, nouns/verbs, word form, and word ending. Errors involving the first six features interfere with the reader's comprehension; however, errors involving the last two features seem to cause the reader less difficulty in understanding the intent of the writer.

Prepositions. The following is an example of a word choice error in a preposition. A native speaker probably would have used "in."

This celebration is celebrated on the middle of June.

Connectors. Connectors are co-ordinating conjunctions, subordinating conjunctions, relative pronouns introducing a clause, cor-relative conjunctions, and conjunctive adverbs. Note the omission of a co-ordinating conjunction in the following example:

> We would be better off without it because it is very expensive, fragile, [and] affects other sattelites in orbit.

Articles. The following examples shows an ommission of the article "a" in front of "higher."

> I have come from Pakistan to get higher education.

Modifier. Besides regular adjectives and adverbs, I included other parts of speech used as adjectives and adverbs under the feature of "modifiers." The following example shows a word choice error using an adjective. The writer has included it where a native speaker would not.

> These committee is to ensure that enough financial support is sufficient for this project.

Auxiliary. The following example reveals a word choice error in the use of an auxiliary. The writer intends conjecture. A native speaker probably would have used "would" instead of "will."

> During the first invention of laser beam, people never thought that these invention will lead to such tremendous improvement.

Noun/Verb. The noun/verb feature has two major parts of speech included because they are found in all basic English sentence patterns. Perhaps a writer's vocabulary strength or weakness is most easily seen in his choices of the words that carry most of the meaning in an English sentence. If these features were combined into one feature and studied, it might yield meaningful results. "Verbs" in this category, however, do not include errors in tense or

errors in verb ending. The following example is a word choice error using an incompete verb form. A native speaker would have used "to be" instead of "to" in front of the adjective "thankful." Probably the writer mistook "thankful" for a verb and tried to make the form parallel with a later verb phrase, "to remind." The example also shows an ommission of the pronoun "themselves."

> To thankful for and to remind for all the kindness that the natives Indian have given to them, they began to celebrate the thanksgiving day each year.

Another error in pronoun word choice is seen in the following example where a native speaker would probably omit "they":

> The main purpose when the engineers invented robots is that the robots they would leave the world better off.

Word Form. Word Form errors indicate the student's partial knowledge of rules governing affixes and base words. Perhaps a student knows the base adjective "clear" but, for example, has not yet learned what kind of affix to attach to make that word an adverb. Errors in infinitive forms are also included in this feature category. These kinds of errors do not seem to be strictly vocabulary problems. Certainly they interfere with reader comprehension, but they are not as confusing as a completely incorrect word choice would be. It is not so difficult for a native speaker to mentally add the proper affix, infinitive form, or other necessary form and continue on with his reading. Therefore, for this study I separated these kinds of errors from the

other word choice errors to better isolate what the learner yet needs to master.

For instance, the writer of the following sentence used "friendship" where he meant "friendly." He is not showing so much an error in choice of vocabulary word as he is showing partial knowledge of the proper word form.

> United States is remain friendship with the strong nations.

In a second example, the writer used "destruct" when he means "destroy." Yet the reader has little difficulty in mentally substituting the correct word; the communication is not seriously disturbed.

> Also these highly concentrated beam can be used to destruct the spy or enemy satellite.

Word Form Endings. The last feature in the Vocabulary category is Word Form errors in word endings. For this study, Word Form Ending errors involve a word form error ending in -ed or -ing, often verbs or verbals. Not included in the Word Form Ending feature are main verbs not requiring an auxiliary with a simple past or present progressive error ending in -ed or -ing. These main verbs are listed under the Verb feature either because of an error in tense or an omission of an auxiliary. (For example, "I was walk" is a word form error--"walk" is the right word but it has the wrong ending. However, "I walking" is an auxiliary error--the writer has ommited the auxiliary; "walking" is not a word form error since "walking" would be the right

word with the right ending.) Shaughnessy (1977) observes that raters must study the error in context in order to best interpret what category of error it is.

The writer of the following example is using a verb form requiring an auxiliary. He has used the auxiliary correctly, but has the wrong word form for the main verb--in this case an error in an -ed word ending.

> There are some degree of percentage of failure in using the laser of medical purposes, but at least it has help many lifes.

Mechanics Category

Mechanics in this study defines a category most English teachers recognize as containing the features of punctuation and spelling. I include these features here, too, but also I include separate features for the comma (does not include comma splices) and capitalization. My punctuation feature contains all other punctuation errors except commas. In addition, I have included run-on sentences and fragments here as Mechanics features

Run On. Run-on sentences are either two sentences run together without the benefit of a period or other end punctuation, or loosely-related sentences connected only by a conjunction, or two closely-related sentences connected only by a comma, or two closely-related sentences connected by a comma and a conjunctive adverb. In this category,

these errors merely indicate problems with using appropriate sentence boundary punctuation; they do not show a lack of sentence boundary knowledge (see Shaughnessy 1977).

For example, the writer often strings short sentences together with commas. In the following sentence, the writer is aware of sentence boundaries but unaware of correct boundary punctuation:

> Everybody is affected by this problem, the size of it is enormous, it does not respect boundaries.

The following writer used "and" as a sentence ending instead of a semicolon or a period. The two ideas are somewhat related, but not so much that they should occur as one thought:

> An example of that is acid rain in Canada which originated in the U.S. and similar problems exist in Europe.

Fragment. Like the run-on sentence, the fragment is included as a Mechanics feature. Sentences without a subject or a verb or which stand as a single subordinate clause are fragments. In this case, ideas that should be connected are in fact disconnected from each other by a period. The writer seems to have a problem with sentence boundaries but not in the sense that he indiscriminately places a period just anywhere in the sentence. Almost always he places a period after a clause, particularly one that begins with "which." He always seems to have a boundary in mind, but he has not learned the rules for appropriate punctuation between boundaries. More to the point, he has

not refined his categories of boundaries to which he can then apply the appropriate punctuation (see Shaughnessy 1977). In the following two examples the writer mistakenly puts a period where he should have used a comma:

> Since United States is a strong nation. The technology is so advance.
>
> If the war happen again. The world will be in great trouble.

Spelling. In addition to misspelled words, the Spelling feature in the Mechanics category includes words showing incorrect usage of hyphens (For example, "People need to up grade their standard of living."). Homonym errors are also listed in this feature (For example, "People think only in there interest.") British spelling, on the other hand, is acceptable--such as practise, learnt, and modernisation.

Cap, Punct, Comma. Capitalization, punctuation, and comma features are included as separate features in the Mechanics category. Faulty conventional usage is listed under these features. "California" should have been capitalized in the following example:

> If we want to go to california fro Oklahoma City probably it will take two days to get there.

Verb Category

The features contained in the Verb category are errors in subject/verb agreement and tense. Errors in verb endings or forms other than that which could be attributed to agreement or tense are listed under the appropriate Word Form feature in the Vocabulary category. If the main verb is in the correct tense, but the auxiliary is not, I have given the writer credit for knowing "tense," but charged him with not knowing complement auxiliary forms. Therefore, that kind of error is listed under the Word Choice/Auxiliary feature. Only if the auxiliary is clearly the proper word choice, but in the wrong tense, is the error charged to verb tense. The verb endings -ed and -ing errors, wherever they occur--as modifiers, nouns, or verbs, after an auxiliary or as the second verb in a compound, are studied to see if they are word form problems or verb tense problems. The error is always studied in context to determine the intent of the writer. Errors in tense are recorded where the intent of the writer seems clear. Infinitive errors are always word form errors because infinitives do not congugate. Even if the writer conjugates the infinitive in what could be deemed a correct "tense," it is an error in form.

Agr S/V. The following example shows an error in subject/verb agreement. The writer adds an auxiliary where a native speaker would not, a word choice/auxiliary error;

but the main verb "involves" does not agree with its subject "devices."

> Its earthbound controlling devices are also involves the latest technology.

The following second example is clearly a subject/verb agreement error:

> The equipment on the space telescope are more fragile than expected.

Yet, the following third example is not so clear. A form of the verb phrase "to helping" ends in -ing, but it seems the writer intends it to be an infinitive. The writer is not having a problem with the word ending as much as he does not know that infinitives do not conjugate. This is a Word Form problem, not a Verb category feature:

> I am trying to helping him.

Tense. The following example is an error in verb tense. From the context of the rest of the essay, the writer should have used the past tense (caused) instead of the present tense (cause):

> The Oklahoma State University cause a big impact of change in my life.

Noun/Pronoun Category

The last category in this study is Noun/Pronoun. It contains the features of unclear pronoun reference, pronoun agreement, possessive, and number errors.

UnclrPro. By unclear pronoun reference, I mean that the pronoun--"that," "this," "it," for example--does not clearly refer to an earlier noun. Note "they" in the following sentence:

> For example, electricians and plumbers working on cleaned apartments, which means that they may have to be clean again.

AgrPro. Errors in pronoun agreement occur when the pronoun referring to a specific noun or pronoun does not agree with it in case, number and gender.

Everyone has their own customs.

Number. Those errors occuring when showing plurality on nouns are listed under the number feature. In the following example, the writer used "frustrations," plural, instead of "frustruation," singular.

I will be prone to make more error's in my work which will lead to greater and greater frustruations and anxiety on my part.

Possession. Errors in possession occur when the writer neglects to indicate possession on a noun or a pronoun or misuses it. Often this problem occurs with "its/it's." The writer mistakenly used "it's" as the possessive form of the pronoun. For instance, in the above example the writer has written "error's," possessive, when he means "errors," plural.

Discourse

The second level of analysis I conducted was discourse. Raimes (1985), in her study of the composing skills of unskilled ESL students, said, "We must be cautious about letting a pedagogical shift in the teaching of L1 writing determine what we look for in ESL research" (p. 232). She cautions that we should not treat second language students like native speakers but, instead, find out what characterizes them as writers grappling with both a written code and a linguistic code still being acquired. However, my study is concerned with what ESL students produce for a proficiency test. Some of that observation must necessarily include the kinds of features raters feel are important in discourse. Therefore, I decided to use the scoring guides (see APPENDIX A) provided to the team of university raters who first holistically scored the examinations.

From the scoring guides, I selected six discourse features the raters were asked to evaluate: Introduction, Body, Supporting Details, Connective Elements, Organization, and Conclusion. I added one more discourse feature to study: Thesis Statement. The thesis statement is a feature often prominently discussed in American writing texts such as those used in the Oklahoma State University writing program. Though the thesis statment is often included in the introduction, it is separated as a distinctive feature in most writing classes. I wanted to see if raters look for

the thesis statement as a cue to understanding the message in the writing sample. These seven features serve as the guide for my discourse analysis.

The definitions of the discourse features in my study were not necessarily inferred from the scoring guides. These features were defined, as Cooper (1977) suggests, from reading a large body of professional and student academic writing. I wanted to compare my feature definitions to those undefined impressionistic ones of the original holistic raters. My definitions were refined after a pilot study was conducted on 14 of the writing samples.

While the proficiency raters holistically scored both discourse and sentence errors on a scale of 1 to 6, with 6 being the highest, in my study I set a less discriminating scale of 0 to 2 and scored discourse only. I wanted to lessen the subjectivity of my scoring scale by making it more quantitative and less qualitative. I was interested in reducing my personal judgment on the relative quality of each feature. A score of two was given if the observed feature met all qualifications, 1 was given if the feature met the qualifications somewhat, and 0 was given if the feature was entirely absent or did not meet one or any of the qualifications for that feature.

Introduction. The Introduction provides background and leads the reader smoothly into the assigned topic. If the introduction were absent or provided background for

something other than the assigned topic, it would receive a score of 0. If the introduction were brief (one or two sentences) but addressed the assigned topic, it would receive a score of 1. If the introduction functioned as defined, that is, both providing background and leading the reader smoothly into the assigned topic, it would receive a score of 2.

The following example is from a passing exam. The writer is addressing an assigned topic concerned with choosing a problem to discuss and proposing a solution:

> With the current economical problems of Oklahoma State University, many of its departments may not be taking the correct measures to control or even eliminate the reasons for the wrong use of monetary funds (thesis statement). That is the case of the department of Student Services Maintenance. Government regulations, improvisation, lack of capacity of reaction to changes, and need for better management tools, could be the facts that cause the waste of money. The fact that this happens in a time of economical crisis, represents not only a problem but a shame (introduction).

This writer's thesis statement comes as the first sentence. His introduction, which follows, clarifies the thesis statement, leads into the discussion with salient points outlined and addresses the assigned topic, and tells how the student feels about it. The writer's introduction score is a 2.

The second sample introduction comes from a passing exam addressing an assigned topic concerned with a world-wide problem that must be solved in the twentieth century:

> There are many kind of world-wide problems that I believe must be solved by the end of the twentieth

> century. The problem tht I would like to discuss here is about the overpopulation, what I mean by over population is that the world is over populated where, for example, there are more people that can cause not enough space to live, food to eat and other basic things that humen needs. Usually this problem occur at third world country or under development country, such as India and Indonesia.

The writer's introduction leads into the assignment somewhat, perhaps not as smoothly as one would wish, and provides a bit of background by defining overpopulation and delineating it as belonging to poorer, underdeveloped countries. This writer's introduction met the feature's definition "somewhat" and received a score of 1.

Another example of an introduction comes from a failing exam written on an assigned topic which asks the student to discuss a technological invention or development we would be better off without:

> The invention of aeroplane was one of the greatest transportation in the world (thesis statement). It helps to get people feel closer to each other from city to city, state to state and country to country (part of the assignment which asks the student to tell the purpose of the invention).

Since the writer begins his essay directly with the thesis statement and finishes the first paragraph and the rest of his essay with part of the requirements of the assignment, no credit for an introduction is given. The writer's score for the introduction feature is 0.

Thesis Statement. The thesis statement must contain the main message of the essay, be clear (that is, the reader must understand what it says), specifically address the

assigned topic, be placed in a strategically useful position in the essay to advance the message, and tell how the writer feels about his statement. The thesis statement does not have to be one sentence, but may be a composite of adjacent sentences.

The following example was taken from a failing exam addressing an assigned topic concerned with a problem that must be solved by the twentieth century, arguing for its importance and discussing the consequences of allowing it to remained unsolved:

> In today's world, environmental pollution is a very large problem which has been given little importance by the nations involve, mainly, in name of progress (thesis statement). This problem, which is usually endorosed by big industry and goverments, it is seldom viewed as a serious threat by the public because it is hard to detect with the naked eye. This esseay will describe this problem and how it affects the air and water, and how this two affect us.

This thesis statement actually incorporates several sequential sentences to help it address all aspects of the assignment--the reasons, the consequences, the writer's feeling about it. His score was 2.

The second sample is also from a failing exam that addresses a topic which asks him to argue that attending college is or is not worth the investment, and asks for an explanation of the writer's present circumstances:

> So, if we have enough courage, all of the investment that we have made before will be more worthy to be use in the future.

The writer's introduction makes three points: College takes time, investment, and courage. His thesis statement

seems to address the topic, but is not written well enough for the reader to clearly understand what the writer intends to state. Does he mean by "worthy" that he believes the investment is "worth" it? Does he mean that courage will make the investment worthy? Does he mean the investment will be made more worthy when the knowledge he has gained in college is used after graduation? Because the thesis statement seemss critical for the reader's understanding of the intent of the writer, clarity is essential. This statement lacks clarity. The student's score was 0 for this feature.

The final example is from a passing exam assigned to discuss a trend that has been increasing or decreasing over a period of time. The writer is to describe the trend and explain it:

> For about a decade the world of the free, modern, industrial world, it include the United States, have not born enough children to reproduce themselve over an extended period of time.

This writer addresses the assigned topic by declaring he will discuss the current decreasing birth rate in the richer, more developed countries. Earlier in his introduction he said "it will change the United States and the world in which we live." Nonetheless, he doesn't tell whether that change will be good or bad; in other words, he doesn't clearly tell how he feels about his thesis statement (a factor not asked for in the assigned topic, but required by the feature definition). The second paragraph generally refers to the way the writer feels about his thesis and

discusses briefly the description and the explanation of the trend, but it is too far from the thesis statement to help it clearly. This writer received a score of 1 for his somewhat unclear thesis statement.

Body. There are two kinds of Body features: one that describes Body in Essay A and a different one that describes Body in Essay B. Essay A and Essay B differ in writing pattern. Essay A is an argumentative pattern, requiring that the writer form an opinion for or against something and support that opinion with facts, details, examples, illustrations, etc. Generally, the writer in this kind of writing pattern continuously considers his reader in his argument and discusses reasons and consequences of his topic to help convince the reader of the truth of his thesis statement. It is formal, rather than informal; objective, rather than personal. Essay B is a descriptive/explanation pattern. It requires the reader to pick a topic he is personally familiar with. He is to describe something from his personal viewpoint and then to explain something about it. Therefore, the writer must make a sophisticated shift in his writing pattern, from description to explanation. It is informal, rather than formal; subjective rather than objective.

Essay A's Body must be in a rhetorical pattern useful in advancing an argument for or against something. The argument, the reasons, explanations, consequences, and

pertinent points must convincingly support and advance the thesis statement and address all aspects of the assigned topic. Essay B's Body must be in the description/explanation pattern, clearly show a shift from description to explanation, and address all aspects of the assignment.

Perhaps those who received a score of 0 on this feature in Essay A do so because they neglect to address all aspects of the assignment. Perhaps the argument does not support the thesis statement, but instead strays from the original topic, even though it still shows ability to write clearly. If the writer receives a score of 1, he probably addresses all the aspects of the topic assignment, but is weak in supporting the main points. If the writer receives a score of 2, he addresses all aspects of the assignment and his support is strong for his thesis statement.

Those who receive a 0 on this feature in Essay B perhaps do not make a clear shift from description to explanation. The writer might either describe but does not explain or explains but did not describe. If he receives a score of 1, he addresses all aspects of the assignment but does not describe nor explain clearly or well, or his description/explanation is very brief or noticeably uneven. That is, he might have described very well but explained poorly. If he receives a score of 2, he addresses all aspects of the asssignment and clearly shifts from

description to explanation. He describes and explains equally well.

Supporting Details. Supporting details are those that are pertinent to the topic or argument; and that support, clarify, and expand the main points of the argument or topic; and that are in the form of details, illustrations, examples, facts, and figures.

An essay could receive a score of 0 on Body because it does not address the assigned topic. Yet the same essay could receive a score of 2 on the Supporting Detail feature because the details meet all the requirements of that feature even though it might be supporting an incorrectly addressed argument.

A writer receiving a score of two on this feature would meet the definition of supporting details in every way. He need not exhibit every kind of detail but would have obvious, pertinent, and often critical support of the main points. Sometimes the writer might provide critical support for two of his three points, and less pertinent support for one. A score of one on this feature would mean that the writer provides some supporting details, but that they are less pertinent or too few to be useful to the expansion of the main points. He might give one-word descriptions, vague or general examples, or one brief example or illustration on only 1 of his main points. A writer receiving a score of 0

on this feature would give no supporting details which expand or clarify the main points of his argument or topic.

Connective Elements. Connective elements--either one-word, phrasal, or paragraphical--are used to lead the reader smoothly from point to point. Pronouns are seldom used as key connecting words.

Examples of connective elements from student samples are "first," "second," "third," "finally," "however,""on the other hand," "in years past," today," all in all," or "in conclusion." In addition, the writer can take a key word or phrase from one sentence or paragraph and carry it over to the second. The writer is not confused about the connective element incorporated into the definition of a paragraph. That is, he understands that the paragraph has one idea, and that each succeeding sentence flows smoothly into the next because there is a connective element in the ideas within that paragraph. Connective elements help the reader keep in constant touch with the flow of ideas from the writer. The word "smoothly" in the definition of this feature is important. A "smooth" connector is one that is pertinent, useful, does not mislead and otherwise functions properly to enhance the reader's understanding of the discourse.

A writer providing a generous supply of connectors within paragraphs, but always between main points and paragraphs, would receive a score of 2 on this feature. If he provides some useful connectors within his essay, he

would receive a score of 1. But if no useful connectors were used, or the connectors used were misleading, or if the writer had trouble with paragraphing in general, his score would be 0.

Organization. The writer uses a clear, recognizable pattern of organization. That is, the writer orders his material by using a writing pattern an American reader recognizes as useful: narrative, chronological, comparison/contrast, descriptive, explanatory, assignment directed, etc.

Those patterns that aren't acceptable are ones which cause the reader to be confused as to the direction of the message; or perhaps the essay is so short as to have only one paragraph, making it difficult to recognize a distinct organizational pattern. The introduction and conclusion, though expected in most organized writing patterns, are not included in this feature since they are analyzed as separate features. Therefore, for the purposes of Organization, only the body is analyzed.

An essay receiving a score of 2 on this feature would exhibit a clear, recognizable pattern of organization. Most of the time this means that the writer is conscious of the constraints of the assigned topic question and orders his material to suit the question. If he were writing Essay B, for instance, he would organize one part of his writing by description and the second part by explanation. If he were

responding to a question for Essay A, he might, for example, first describe the problem, then propose a solution, and finally explain how his solution would be better than the current situation.

If the essay received a score of 1, the writer might have separated the description from the explanation in Essay B, but kept it all in one paragraph; or, he might have begun the description, shifted to the explanation, and without an obvious reason returned to the description. In these cases, the writer shows an understanding of organization, but not clearly so. The reader must adjust his interpretation to understand the message, a function good organization would have provided for him otherwise.

An essay receiving a score of 0 had no observable organization. The writer might shift from one point to another and back again. He might place the discussion of one point before another, the second being crucial to understanding the first. His material might be so vague or so general that the reader cannot see that he has made any points at all and, therefore, is unable to see a pattern to the message.

Conclusion. The final discourse feature is the Conclusion. This feature is defined as summarizing the main points of the discourse and/or projecting to the reader questions or future considerations about the issues already discussed there.

The essay containing a conclusion summarizing the main points, using different words or phrases from those used in the body, is given a score of 2. In this case, the writer doesn't simply restate in almost the very same way the points supporting his thesis statement. He might conclude his essay with a projection into the future, asking hypothetical questions or proposing hypothetical situations concerning his thesis statement. Note the following example of a conclusion rated 2 from a failing paper:

> I strongly urge the public as well the media to oppose mandatory testing for certain groups of people because it deprives them their freedom; brings more misery in their lives, and creates unwarranted descrimination. This is like telling some one that you are going to die withought helping him or her to live. Money should be spent on researching a vaccine for Aid virus not a mandatory testing of selected individuals.

This writer is exhorting someone to do something. He also lists the main points again and makes recommendations. It meets the requirements for the feature Conclusion.

The essay giving a conclusion of only one or two lines, even if it otherwise generally followed the feature definition, received a score of 1. In this case the conclusion is not developed enough to allow the reader to feel he has grasped the message in a capsulated form. The following example from a passing exam received a score of 1 on this feature:

> Overall, I hope I could have a beautiful life and a bright future.

The writer's conclusion vaguely fits the part of the feature definition about projecting into the future, though

the reader is not left with a feeling that he has the essay's message in summary.

The following is another example of a conclusion, rated 1:

> Remember that this problem only can be solved if there is a cooperation and, trust exist.
> Finally, remember that this problem can give bad consequences to individual life, if we never try to solve it.

The writer of this conclusion is projecting consequences into the future in a general sort of way. However, the reader is not left with the feeling that he has a summary of the message of the essay.

And a third example of a conclusion from a passing exam with a 1 rating:

> Over the past twelve years, many people have had their names written in history books for their contribution in destroying Lebanon. I like my name to be engraved in as the person who helped rebuilding it.

This writer left the reader with a feeling that he has just reread the exact thesis statement, which he had. The main points of the essay are not summarized (Note: In truth, it was a personal essay and might not have lent itself easily to that); but an idea of the future is projected, yet it only vaguely fits the requirements of the conclusion feature.

The essay with no conclusion or with a conclusion of one line which did not fit the definition of this feature received a score of 0.

The following example of a conclusion rated 0 is from a failing exam:

> For those reasons and many others, I think that the first time when the men walked on the moon was a great event in the human history.

The writer does not summarize the reasons, nor does he explain the "many others," and he simply repeated his topic sentence. This "concluding" sentence was added on to the end of a paragraph discussing the last reason explaining his thesis statement. It is not even clear whether or not the writer intended this sentence to be the conclusion or if he meant for it to accompany his last paragraph. If the latter is true, then he wrote no conclusion and, therefore, would also have received a score of 0 for that feature.

Procedure

After the features to be analysed were categorized and defined, I conducted a pilot study for each level of analysis to refine the descriptors.

In the sentence error analysis, I marked seven passing samples and seven failing samples--28 writing tasks in all--according to the categories and features first outlined. In this way, I was able to study the writing in context and look for common patterns of faulty logic which might be systematic and therefore important to my study of second language learners. I was not interested in using an error taxonomy refined from the writing of native speakers and simply generalizing it to the writing found in my study of non-native writers. Therefore, the contextual study

helped me to clarify each category and which features it would be limited to; in addition, I felt my categories represented what the non-native writers in my study were actually writing. I finally refined the sentence error analysis taxonomy into 5 separate categories subdivided into a total of 22 features.

I conducted a second pilot study to refine the discourse features I wanted to analyze. Fourteen examinations were used to clarify the discourse feature descriptions outlined in my study. In this pilot study, I was particularly aware of the general description of "good rhetoric" as explained on the original holistic scoring guides (see APPENDIX A). Also, I was aware of "good writing" as explained in typical writing texts such as those used in the Oklahoma State University writing program. I was interested to see whether Cummins's CALP as it might apply to classroom academic writing would be revealed in the distinctions of writing proficiency on these examinations. Would raters look for features on proficiency tests that they might have specifically taught in the classroom. For the discourse analysis, all sentence errors and incoherent constructions where they didn't interfere seriously with comprehension were ignored. My discourse taxonomy was finally limited to seven features and its descriptions clarified. After I had completed the pilot study on discourse errors, I scored all the examinations once again at one sitting to help remove the bias that I may have

scored differently from sample to sample and from day to day.

Once the categories and features were refined, the writing samples were marked for errors; and the errors were counted and recorded.

Statistical Discussion

The error analysis and discourse data were examined from a statistical viewpoint. Standard equations were used to obtain total, percent of total, mean, and standard deviation values for each feature. In addition, correlation coefficients (r-scores) were calculated for the features, categories, and for the error analysis and discourse total scores versus the exam scores (proficiency test scores). In addition, the correlation coefficient between the error analysis and discourse score was determined. All calculations were done on an IBM PC using the Symphony program.

Consider two students (A and B) of exactly equal writing abilities. Student A writes an essay of a thousand words while student B writes a five hundred word essay. Since they are of the same ability, writer A, with the longer essay, will obviously have a greater total number of errors when evaluated by error analysis. In fact he would have twice as many errors as writer B since his essay is twice as long. A rater might erroneously conclude that A is

a much poorer writer than B based on the total number of errors. However, if the number of errors for each essay is divided by the number of words in that essay, the result is an error frequency that reveals the two writers to be of equal ability.

The error analysis results were normalized by determining an error frequency based on the number of words written by each group. The normalization was done in the following manner. The total errors for each group were divided by the total number of words written by that group. The result was the frequency in errors per word for the entire passing and failing group. For example, there was a total of 4360 errors recorded on the error analysis and a total of 32539 words written. By dividing the errors by the words we get an error frequency of 0.134 errors per word. Since this number was very small (0.134) the errors per word were multiplied by 1000. This gave a number (134) that was easier to discuss and is by definition the Normalized errors (N-error or N-e).

N-e is the frequency of errors per 1000 words. It is the number of errors that occur in one thousand words. As another example, the passing group had a total of 1911 errors and had written a total of 18656 words. This corresponds to an N-error of 102.4 for the passing group. The failing group had 2507 total errors and had written 13883 words. Their N-error was 180.2.

In addition to determining N-e for the total error analysis, N-e's were determined for each category and feature by group, either passing or failing. For example, in the Vocabulary category, the passing group made a total of 736 errors while the failing group made 971 errors. N-e for the passing group is 39.5 (736 errors divided by 18656 words) and 69.9 for the failing group (971 errors divided by 13883 words).

The same argument of length is not necessarily true for the discourse results. A pithy response may well cover all the needed points while a longer essay may never come to the point. On the other hand, a short response may indicate lack of knowledge or writing skill and a long essay may indicate confidence, organization, and a high level of proficiency. Therefore the discourse scores were not normalized for length. However, since there were 20 students in the passing group and 19 in the failing group, the mean score was used instead of the total score for all comparisons between the two groups. For example, the passing group made a total of 327 points on the discourse analysis. There were 20 students in the passing group. The mean passing score was, therefore, 16.3 (327 points divided by 20 students). The mean total score for the failing group was 10.3 (195 points divided by 19 students).

In order to see differences between the two groups more clearly, the ratio (R) of the failing group's N-error score to the passing group's N-error score was used when

discussing the error analysis results. For example, in the Vocabulary category the failing N-e was 69.9 and the passing N-e was 39.5. This gives a ratio (R) of 1.8 (69.9 divided by 39.5). From this number we can see that the failing group made 1.8 times as many errors as the passing group. Ratio is also used in the analysis of the discourse results to indicate the relative abilitiy of the two groups. For the discourse results, however, ratio (R) is the mean passing score divided by the mean failing score. Using the total discourse mean scores, R is 1.7 (16.3 divided by 10.3). This means that the passing group received 1.7 times as many discourse points as the failing group. Both of these ratios reveal the relative difficulty the failing group had with a particular feature compared to the passing group .

It should be kept in mind that the two groups were determined by the original proficiency exams. The groups were <u>not</u> reordered based on the results of the error and discourse analyses. The original groups were preserved so that the differences determined by the detailed analyses might be related to the holistic exam scores.

CHAPTER IV

RESULTS, DISCUSSION AND IMPLICATIONS

Introduction

This chapter contains the results of the analysis of the performance of thirty-nine international university students who took the English writing proficiency test at Oklahoma State University. A total of seventy-eight writing samples were analyzed on two different levels of writing after initially being scored holistically by a team of university raters. First, the writing samples were analyzed on the sentence level for surface errors in five categories of features. Next, they were analyzed at the discourse level for performance in seven features.

The data were analyzed with regard to the following questions:

1. Are there features between those who pass and those who fail which might clearly determine proficiency?
2. What kinds of distinctions between those who pass and those who fail would a sentence error analysis and a discourse analysis show that a holistic examination score would not show?

3. How do the sentence error and discourse analyses scores compare with each other and with holistic scoring as a test of general proficiency?

4. What do proficiency raters seem to attend to when judging proficiency?

Results

Error Analysis

The error analysis consisted of twenty-two features divided into five categories. For each feature the total number of errors was counted and normalized. In all cases where applicable, normalized errors (N-errors) were discussed instead of the actual errors (TotErr). Then, the mean, range, standard deviation, total normalized errors, and the percent of the total normalized errors were calculated. Next, for each of the seventy-eight writing samples the number of words and errors was counted, and the normalized errors per word were calculated. All correlations were calculated against the original holistic examination score, hereafter called "exam score." Finally, the five categories of features were subjected to the same statistical analysis, with the addition of the correlation coefficient and the ratio.

The data were further separated between passing (P) and failing (F) scores as indicated by their holistic exam score

and separated by writing task. The writing task requiring persuasion was identified as task "A", and the writing task requiring description/explanation was desiginated task "B."

Of the statistics generated, three indicators were found to be particularly useful in examining the results: the correlation coefficient (r-score), N-errors, and the ratio (R) of the N-errors. The r-score is an indication of how well a particular category or feature correlates to the exam score. In this study, an r-score greater than .32 is statistically significant. Any r-score greater than .32 may indicate a distinguishing characteristic of the failing group. N-errors measure the frequency of errors and could indicate the seriousness of the error problem. Ratio is the N-errors in the failing group divided by the N-errors in the passing group for a particular feature or category. It indicates the relative frequency of errors in the failing group when compared to the passing group. It could be used to determine distinguishing characteristics of the failing group.

When the correlation coefficient is greater than 0.32 and the ratio is high, a high N-error indicates a significant characteristic of the failing group. If N-error is low, this may indicate an irritant to the rater.

When the correlation coefficient is less than 0.32, the feature is not statistically significant as a determiner of pass or fail. If the ratio is high and N-error is low, the feature is not meaningful, possibly due to the small number

of errors. If the ratio is normal to low and N-error is high, this indicates a major problem for both groups. If the ratio is normal to low and N-error is low, then neither group is having a problem in this area.

When the correlation coefficient is significant ($r > .32$) and the other indicators are high (category R > 1.8 or feature R > 2.0; category N-error > 26.8 or feature N-error > 6.1), then the feature or category could be considered a characteristic of the failing group.

First, the categories of the features were studied statistically in order to determine their relative significance to the three indicators.

I found that passing students wrote 1.3 times as many words as did the failing students (Pmean=933; Fmean=730) and the failing students made 1.3 time more errors than did the passing group (F=2502; P=1911) (see TABLE I, p.70).

First of all, I found that two categories had significant r-scores: Vocabulary (-0.50) and Sentence (-0.40). These two categories show a moderate negative correlation to the exam score. That is, the more errors made in these categories, the lower the exam scores. Both categories also had high ratios of errors between groups (1.8). However, Vocabulary had high N-errors (P=39.5; F-69.9) while Sentence did not (P=7.7; F=13.7). Therefore, I found that Vocabulary could be an important category which might have features which, when examined, could distinguish between those who passed and those who failed. The sentence

category could also be important, but needed further study to determine which features account for its significance.

TABLE I

ERROR ANALYSIS CATEGORIES

	Vocab	N/P	Verb	Sent	Mech	NoWds	TotErr	N-e
PASSING								
TotErr	736	254	187	143	591	18656	1911	
%TotErr	39.6	13.7	10.1	7.7	31.8		43.3	36.2
TotN-e	39.5	13.6	10.0	7.7	31.7			102.4
Mean						933		
FAILING								
TotErr	971	309	228	190	804	13883	2502	
%TotErr	38.8	12.4	9.1	7.6	32.1		56.7	63.8
TotN-e	69.9	22.3	16.4	13.7	57.9			180.2
Mean						730		
Ratio	1.8	1.6	1.6	1.8	1.8	1.3	1.3	1.76
rScore	-0.51*	-0.24	-0.11	-0.40*	-0.20	0.60*	-0.47*	-0.66*

*P < .05

Mechanics had high N-errors (P=31.7; F=57.9) and a high ratio (1.8), but its r-score was lower than .32 (-0.20). The low correlation to the exam scores for this category meant that it was not statistically significant as an indicator of pass or fail.

All the N-errors together had a moderately significant correlation with exam scores (-0.66), and the number of words written correlated moderately well with the exam score (-0.60.). That is, the more words written, the higher the exam score. Because two categories, Vocabulary and Sentence, had moderately significant correlations with exam scores, r-scores were then calculated for each of the features in those two categories to try to determine which features, if any, contributed most to that significance.

Each of the categories was then broken down into its separate features to study.

TABLE II

VOCABULARY

	Prep	Conn	Art	Mod	Aux	N/V	WForm	WFending
PASSING								
TotErr	114	48	233	31	64	109	39	98
%TotErr	6.0	2.5	12.2	1.6	3.3	5.7	2.0	5.2
TotN-e	6.1	2.6	12.5	1.7	3.4	5.8	2.1	5.3
FAILING								
TotErr	147	43	272	65	51	197	89	107
%TotErr	5.8	1.7	10.8	2.6	2.0	7.8	3.5	4.2
TotN-e	10.6	3.1	19.6	4.7	3.7	14.2	6.4	7.7
Ratio	1.7	1.2	1.6	2.8	1.1	2.5	3.0	1.5
rScore	-0.45*	-0.33*	-0.36*	-0.43*	-0.15	-0.49*	-0.44*	-0.12

*P< .05

Six of the eight vocabulary features had r-scores above .32: Noun/Verb (-0.49), Prepositions (-0.45), Word Form (-0.43), Articles (-0.36), and Connectors (-0.33) (see TABLE II). However, only one feature had all three indicators above normal--Noun/Verb. This feature had high N-errors in the failing group (14.2), a high ratio of errors between the groups (2.5), and a low moderate correlation coefficient significance (-0.49). It might be a feature that distinguishes pass from fail. Raters might attend to word choice errors in nouns and verbs because they carry most of the meaning in an English sentence, and statistically significant errors disturbing comprehension might be major irritants to raters.

Though the Article feature was highest in error frequency, it might be relatively unimportant to raters. Its ratio was 1.6--less than some categories with far fewer N-errors--and its r-score indicated only a low correlation to the exam score.

The Preposition, even with a -0.45 significance, may not distinguish between pass or fail since errors have a normal distribution (for this study) between groups.

The Connector feature had low N-errors, low Ratio (1.2) and a low r-score significance (-0.33). Because the number of errors was so low, its significance to raters cannot be assessed.

The Modifier and Word Form features had very high ratios, moderately low r-scores, and very low N-errors. The

N-errors were too low in these two categories to predict rater bias.

TABLE III

NOUN/PRONOUN

	UnclrPro	AgrPro	Poss	Number
PASSING				
TotErr	18	6	20	181
%TotERR	1.0	0.3	1.0	9.5
TotN-e	1.0	0.3	1.1	9.7
FAILING				
TotErr	56	16	11	226
%TotErr	2.2	0.6	0.4	9.0
TotN-e	4.0	1.1	0.8	16.3
Ratio	4.0	3.7	0.7	1.7
r Score	-0.39*	-0.06	0.18	-0.30

*P< .05

The Unclear Pronoun feature (see TABLE III) had a low correlation to the exam scores and a very high ratio of errors between the groups; however, the N-errors were low and, therefore, did not clearly indicate any rater bias.

By far, Number is the feature with the most N-error frequency in this category by both groups (P= 9.7; F= 16.3). However, the high number of N-errors may not be as important to the rater (R=1.7) as other features with higher error

ratios and significant correlations. Interesting to note is that the failing group made fewer N-errors (0.8) in possession than did the passing group (1.1).

TABLE IV

VERB

	AgrS/V	Tense
PASSING		
TotErr	60	106
%TotErr	3.1	5.5
TotN-e	3.2	5.7
FAILING		
TotErr	104	124
%TotErr	4.1	4.9
TotN-e	7.5	8.9
Ratio	2.3	1.6
r Score	-0.09	-0.26

*P< .05

The total Verb category errors, when normalized for each feature, indicated that while more errors occurred in tense in both groups, Subject/Verb Agreement N-errors occured much more frequently in the failing group (7.5) than in the passing group (3.2). Yet, the correlations of these features with exam scores were insignificant in both verb category features.

TABLE V

SENTENCE

	Parallel	IncConst
PASSING		
TotErr	11	115
%TotErr	0.6	6.1
TotN-e	0.6	6.2
FAILING		
TotErr	7	183
%TotErr	0.3	7.3
TotN-e	0.5	13.2
Ratio	0.8	2.1
rScore	0.05	-0.56*

*P < .05

In the Sentence category, Incoherent Constructions indicated moderate significance with exam scores (-0.56). In addition, its ratio (2.1) and its N-errors (F=13.2) were high. In fact, all three indications of rater importance were high. The failing group had much more trouble with Incoherent Construction compared to the passing group. As with word choice errors in nouns and verbs, incoherent construction interferes with reader comprehension. Perhaps raters are sensitive to errors disrupting comprehension. Parallel Construction has no significance when correlated with the exam scores (.05).

TABLE VI

MECHANICS

	Frag	RunOn	Spel	Punct	Comma	Cap
PASSING						
TotErr	39	77	192	57	156	84
%TotErr	2.0	4.0	10.0	3.0	8.2	4.4
TotN-e	2.1	4.1	10.3	3.0	8.4	4.5
FAILING						
TotErr	70	107	257	44	210	116
%TotErr	2.8	4.3	10.3	1.8	8.4	4.6
TotN-e	5.0	7.7	18.5	3.2	15.1	8.3
Ratio	2.4	1.9	1.8	1.1	1.8	1.8
r Score	-0.48*	-0.20	-0.16	0.07	-0.07	-0.09

*P<.05

In the Mechanics category, the feature with the highest r-correlation was Fragments (-0.48) It also had a high ratio of errors (2.4), but low N-errors. This feature may be an irritant to the rater, but the low error frequency did not clearly show it to be a determinant of pass or fail. Spelling had the highest N-error frequency in this category (P=10.3; F=18.5). Nearly equalling spelling in N-errors is the Comma feature (P=8.4; F=15.1). However, both groups made more N-errors in those features than average, and their r-scores indicated they were not significant to the rater.

Discourse

While the error analysis studied five categories of features, the discourse analysis studied seven. The combined discourse scores of each of the features gave an overall view of how well the writers performed (see TABLE VII). Since there were twenty students in the passing group and nineteen in the failing group, it was necessary to use the Mean scores rather than the total scores when making comparisons. The Ratio (R) is the mean passing score divided by the mean failing score. When all indicators are high ($r > .32$; $R > 1.8$; mean > 1.9), then the feature could be considered a characteristic of the passing group.

After I scored the writing samples for discourse, I totaled the scores for each feature and recorded them. The writers earned a total of 91 points for Organization, one of their best features. In this feature, each writer who passed earned an average of 3.00 of those points while each who failed earned 1.63 points. (The mean score was used to compensate for the unequal sample size--20 passing students and 19 failing students.) This is a ratio of 1.8 which means that, on the average, the passing students earned nearly twice as many points in this feature as the failing students. The r-score for Organization indicated it was statistically significant (.50).

TABLE VII

DISCOURSE FEATURE SCORES

	Intro	ThState	Body	Details	Conn	Organ	Concl
PASS TOT	33	48	57	39	52	60	38
%PASS	10.1	14.7	17.4	11.9	15.9	18.3	11.6
MEAN	1.65	2.40	2.85	1.95	2.60	3.00	1.90
FAIL TOT	27	31	26	29	28	31	23
%FAIL	13.8	15.6	13.3	14.9	14.4	15.9	11.8
MEAN	1.42	1.63	1.37	1.53	1.47	1.63	1.61
TOTALS	60	79	83	68	80	91	61
MEAN	1.54	2.03	2.13	1.74	2.05	2.33	1.56
STD. DEV	1.06	0.97	1.22	0.90	1.13	1.25	1.13
Ratio	1.2	1.5	2.1	1.3	1.8	1.8	0.6
r SCORE	0.23	0.44*	0.50*	0.30	0.49*	0.50*	0.31

*P < .05

The Body feature had the next highest average number of points (57). Note that those who passed made over two times as many points per student as those who failed (R=2.1). Furthermore, the r-score was significant (r=.50). In addition to Body and Organization, the passing group did significantly better in Connective Elements (r=.49) and Thesis Statement (r=.44).

In three of the features with significant r-scores and high ratios, the passing group had high mean scores which when taken together might indicate that they were specific characteristics of the passing group. These three features might distinguish pass from fail. Body might be the most

important determining feature. It had an r-score of .50, a ratio of 2.1, and a passing mean of 2.85, all above average. Organization was a second distinguishing feature according to this study (r=.50; R=1.8; P mean=3.0). Connective elements was the third distinguishing feature (r=.49; R=1.8; P mean=2.60). All three features are closely involved with the basic meaning of the rhetoric pattern and statistically significant. Again, it is possible that raters are more irritated by these features than other rhetoric features because they interfere with basic comprehension.

The failing group performed uniformly on all the other features. Unlike the failing group, the passing group's results indicated a wide range of performance on the discourse features.

The results of the discourse totals (see TABLE VIII) showed that the combined number of points scored by the writers was 522. Each writer could earn a possible total of 14 points per task (28 points for each examination). The mean score for all writers was 13.38 points. The passing group scored above the mean (16.35 pps) while the failing group scored lower than the mean (10.26 pps). The discourse scores correlated moderately well (0.65) with the original holistic exam scores.

Of the 522 total points, the passing group earned 327 of them, or 62.6 percent; the failing group earned 195 points, representing 37.4 percent of the discourse total. The percent of the total number of points for each group in

the study (P= 62.6 percent; F= 37.4) was surprisingly close to the percent of the same groups on the holistic exam score (P= 62.0; F= 38.0).

TABLE VIII

DISCOURSE COMPARISONS

	TotDis	ExScore
PASS TOT	327	230
%PASS	62.6	62.0
MEAN	16.35	11.5
FAIL TOT	195	141
%FAIL	37.4	38.0
MEAN	10.26	7.42
COMBINED TOTALS	522	371
MEAN	13.38	9.51
STD.DEV	4.73	3.03
Ratio	1.59	1.55
r SCORE	0.65*	1.00*

*P< .05

Finally, the r-score of the combined total discourse score (TotDis) was compared to the r-score of the combined normalized error analysis score (N-e) (See TABLE IX).

TABLE IX

CORRELATION COEFFICIENTS

	TotDis	TotN-e
r Score*	0.65	-0.66
r Score**	-0.50	

*correlated to the exam score
**correlated to the normalized errors
P < .05

Both the study's discourse score and its error analysis score indicated almost equal correlations with the exam score. From these results it appeared that discourse was given no more weight than surface errors in evaluating the proficiency of these writing samples. In addition, the correlations were moderately high. That is, the higher the discourse score, the higher the exam score; conversely, the more errors, the lower the exam score.

The sentence error analysis and discourse analysis correlated moderately with each other (-.50). That is, the lower the discourse score, the higher the number of N-errors.

Discussion

These results led to some tentative conclusions about international students' performance on written proficiency

examinations. The data showed that the failing group made 1.76 times as many errors in all categories as the passing group (see TABLE I, p. 70). In addition, they wrote less well. The failing group earned only 37.4 percent of the discourse points given (see TABLE VIII, p. 80). Furthermore, they didn't write as much. On the average, the failing group wrote 730 words while the passing group wrote 933 words (see TABLE I, p. 70).

However, the first thing I wanted to know was whether there were features between those who passed and those who failed which might clearly determine proficiency; in addition, I wanted to know what raters might be looking for in scoring the examinations of non-native speakers. Of the five categories studied in the error analysis, all non-native students seemed to have trouble with vocabulary and mechanics. However, no one category seemed to clearly stand out as being significant in determining the difference between passing and failing.

Nevertheless, though distinctions may not be clear when comparing categories of errors, some of the features within those categories in the error analysis might be significant in indicating the difference between passing and failing. If we examine the error analysis data using the ratio as a criterion, we can divide the data into three regions. The first region is $R < 1.5$; the second region is $1.5 < R < 2.0$; the third region is $R > 2.0$. A second criterion we can use to determine the relative importance of the feature is the

average N-errors per feature of 6.1 (see TABLE X). The third criterion indates the significance of each feature when correlated with the exam score.

The average number of N-errors per feature is 6.1. Looking first at R < 1.5, we see that no features were statistically significant. Only one feature had an N-error greater than the average (Word Form -ing -ed). The low ratios (.7 to 1.5) indicated a somewhat random distribution of these errors between passing and failing. These six low ratio features represented 27 percent of the total features but only 10.5 percent of the failing N-errors and 16 percent of the passing N-errors. Apparently, these six features were not important to the raters in determining passing and failing.

Second, looking at the eight features with R from 1.5 to 2.0, we see the bulk of the errors for both passing and failing groups. The ratio indicated a usual distribution of errors between passing and failing. The eight features represented 36 percent of the total features but 58 percent of the failing N-errors and 61 percent of the passing N-errors. Therefore, the majority of the errors for both groups occurred in these eight features. The most troublesome was Articles, followed by Spelling and Number. This is the region where changes in instruction can be used to improve the overall performance of non-native students on proficiency tests. However, there was little evidence of rater bias in the data. Only two features appeared

statistically significant, Articles (r=-0.36) and Prepositions (r=-0.45). Yet, these features show only normal distribution of errors between the groups. These two features do not appear to clearly distinguish those who pass from those who fail.

TABLE X

RANKED ERROR ANALYSIS FEATURES

Feature	Ratio	N-errP	N-errF	r Score
R < 1.5				
Poss	0.7	1.1	0.8	0.18
PllConst	0.8	.6	0.5	0.05
WC Aux	1.1	3.4	3.7	-0.15
Punct	1.1	3.0	3.2	0.07
WC Conn	1.2	2.6	3.1	-0.33
WF-ing-ed	1.5	5.3	7.7	-0.12
1.5 < R < 2.0				
WC Art	1.6	12.5	19.6	-0.36*
Tense	1.6	5.7	8.9	-0.26
WC Prep	1.7	6.1	10.6	-0.45*
Number	1.7	9.7	16.3	-0.30
Spelling	1.8	10.3	18.5	-0.16
Comma	1.8	8.4	15.1	-0.07
Cap	1.8	4.5	8.3	-0.09
RunOn	1.9	4.1	7.7	-0.20
R > 2.0				
IncConst	2.1	6.2	13.2	-0.56*
Agr S/V	2.3	3.2	7.5	-0.09
Frag	2.4	2.1	5.0	-0.48*
WC N/V	2.5	5.8	14.2	-0.49*
WC Mod	2.8	1.7	4.7	-0.43*
Agr Pro	3.7	.3	1.1	-0.06
UnclrPro	4.0	1.0	4.0	-0.39*

*P < 0.05

However, the final region, R > 2.0, might very well indicate significant distinctions between the passing and failing group as well as reveal some possible irritants to the raters. The two significant features were Noun/Verb (N/V) and Incoherent Construction (IncConst). Both had high N-errors and high ratios. Both of these features are associated with reader comprehension. Nouns and verbs carry most of the meaning in an English sentence. If the writer is having trouble choosing the correct noun or verb for his sentence, the meaning might be distorted. Obviously, if the sentence structure is incoherent, it adversely affects the meaning. Perhaps the rater is most irritated by factors which interfere with comprehension. If so, these two features indicated that they were an important source of errors for the failing group. Furthermore, they might reveal a significant difference between them and the passing group. Other possible bias or less significant irritant factors for the rater seem to be errors in Unclear Pronouns (UnclrPro) (r=-0.39), Modifiers (Mod) (r=-0.43), and sentence Fragments (Frag) (r=-0.48) since all have low N-errors, but high ratios.

The discourse analysis data also seemed to indicate a distinction between those who pass and those who fail. Three features were significant among the passing group: Body, Organization, and Connectors (see TABLE XI). These three features had significant r-scores, ratios above the

average (1.6) and scores significantly above the average points per student (1.9):

TABLE XI

RANKED DISCOURSE FEATURES

Feature	Ratio	Mean P	Mean F	r Score
Conclusion	0.6	1.90	1.61	0.31
Intro.	1.2	1.65	1.42	0.23
Details	1.2	1.95	1.53	0.30
Th.State	1.5	2.40	1.63	0.44*
Conn.	1.8	2.60	1.47	0.49*
Organ.	1.8	3.00	1.63	0.50*
Body	2.1	2.85	1.37	0.50*

*P < 0.05

The failing group wrote less well than the passing group; in fact, they scored uniformly low in all features (see TABLE XI). The passing group, on the other hand, scored significantly higher on Body, Organization, Connectors, and Thesis Statement; they scored least well on Details, Introduction and Conclusion, yet still higher on each feature than the failing group. The three most significant features, Body, Organization, and Connectors probably were the features the raters were attending to when scoring non-native examinations. Perhaps the raters saw these three features as being instrumental for overall

comprehension and, therefore, used them to judge overall proficiency.

Finally, I wanted to know whether this study's scores correlated with the holistic exam scores and with each other as a test of overall writing proficiency.

My analyses scores correlated moderately well with the exam scores (see TABLE IX, P. 81). In addition, both analyses yielded the same correlation and nearly the same percent/pass percent/fail results as did the exam scores (see TABLE XII). All three scoring methods yielded nearly 63 percent of the total points to the passing group and nearly 37 percent of the total points to the failing group. On the other hand, the sentence error and discourse analyses scores correlated only moderately well together(r=0.50).

TABLE XII

COMPARISON OF SCORING METHODS

	ExamSc	ErrAn*	DisSc
%PASS	62.0	63.8	62.6
%FAIL	38.0	36.2	37.4

*Note: Since the error analysis has a negative correlation to the exam score, it was necessary to subtract the percents in TABLE I from 100 percent to make the correlation positive for use in this table.

My study supports Kaplan's (1987) belief that at least two levels of analysis are needed to give meaningful conclusions. Meaningful conclusions would include a list of those systematic errors in sentences and faulty rhetoric patterns in the learner's writing. Bhatia's study (1974) picked out features by error frequency, but we need more than error frequency to determine proficiency and to improve instruction. We need error frequency and error gravity indications. It may be impossible to eliminate all errors from L2, even those which occur frequently. Therefore, we need to determine which of those "frequent errors" when corrected will relieve reader irritation and thereby improve communication.

Implications

This study has several implications for writing instruction and testing programs. If proficiency tests require non-native students to write with a university writing fluency, then we must understand the problems of the student. It is not productive to think of him as a basic or remedial writer; we must study him for what he is, a second language learner. While it may be true that some second language learners are not proficient in their first language, that may not be the case for all of them. Whatever their L1 proficiency, Richards (1974a) indicates they bring it with them into the L2. In L2 they struggle

with the rhetoric even as they grapple with the linguistic code.

This study seems to indicate that by helping the international university student to eliminate word choice errors with nouns and verbs and to correct incoherent construction to effect better meaning, the ESL teacher may be able to improve written communication. In addition, by helping him attend to the meaning in the body of his text, and advance the meaning through the use of appropriate connective elements and American organizational patterns, the teacher may be able to improve reader comprehension. Furthermore, the international student, whether he passed or failed an English proficiency test, still must eliminate many features which indicate serious error gravity.

In addition, it seems the failing international student needs to write more. Those who failed wrote significantly less than those who passed ($r=0.60$). This may be due to a lack of academic vocabulary, as indicated by the large number of errors and significant correlations in the word choice category of Vocabulary. Rollins (1985) found that academic reading and writing correlated well. Perhaps by introducing a large amount of academic reading into the international writing program, the international student would be able to build his academic vocabulary bank faster and thus be able to write closer to university proficiency standards.

If proficiency raters seem to be paying attention to those features which advance the meaning of the message, instruction should encourage those factors which aid the second language learner to develop native-like discourse schemas and thought patterns. This means that universities should encourage native/non-native social interaction. Universities should not always isolate non-native speakers into "special" writing classes; they should encourage an integrated writing program with native speakers.

Attending to that which advances the knowledge of discourse schemas and thought patterns is more than social interaction. Raimes (1987) has found that good writers draw from their own background to write. Therefore, socialization with American students in and out of the classroom will help second language learners to improve their supply of language options--that is, their L2 word choice language bank and rhetoric pattern preferences.

Salient feedback on writing samples would aid the ESL student in understanding what is expected of him. Since raters on proficiency tests seem to reward what advances meaning and penalize what interferes with it, the instructor should not mark what is mechanically wrong with the writing at first if it does not disturb comprehension. Rather, she should focus on telling the student what interferes with meaning--both for the sentence and the discourse as a whole. Later, once the student is motivated to attend to his audience and purpose, less significant features could be

marked for revision. Recent research with revision and salient feedback use the cognitive process developed by Flower and Hayes (1981). In this method, the instructor talks with the student, and together they determine what is wrong with comprehension. Shaughnessy (1977) also advocates working closely with the student and his writing to understand why errors are occuring in comprehension.

Holistic scoring yields generally good results when scoring quickly as for proficiency testing. However, if a more thorough analysis is needed, a combination of an error analysis and a discourse analysis might distinguish significant individual factors which, when properly attended to by the instructor, might cause the writer to be more proficient in his writing.

An ESL student may never be in such an advantageous situation again which would create the motivation necessary to master his "deviant language." If classroom teachers and testing programs could capitalize on the tools necessary to create improvement in writing and the motivation of the student, they might be able eliminate that in the international student's writing which causes raters to judge him not proficient. Thus, he might be better prepared to compete with native speakers on a test of English writing proficiency.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

Summary

The Problem

Unlike speaking, writing is a learned skill. It is neither built into our genetic code, nor is it universally distributed among the cultures of the world. Therefore, a speaker may or may not have learned to write his native language well, if at all; and what he has learned might be peculiar to his culture.

From L1, a second language learner brings into L2 all his writing capabilities, strategies, and thought patterns. In addition, he struggles with a new linguistic code. At American universities, he is introduced to unfamiliar social, cultural, and academic situations which require new thought patterns, vocabulary banks, and writing strategies. Communication suffers. More often than not, his writing is judged not proficient on tests of writing skill when he is compared to native writers. Often, further English language instruction at the university is not enough to help him achieve the standards of university writing proficiency

needed to pass required writing proficiency tests.

The objective of this study has been to determine whether measureable distinctions exist between passing and failing groups of non-natives who took an English writing proficiency examination and how that information might affect ESL instruction and writing testing programs. To this end, the data were examined to discover whether distinctions could be found in L2 writing which identify it as proficient or non-proficient. Instrumental in such a proficiency judgment is what raters seemed to attend to and whether the scoring method made a difference.

Sample, Features, and Procedure

The sample in this study was randomly chosen from 99 second language students who took a writing proficiency examination required of all undergraduates at Oklahoma State University during an eleven month period from May 1987 to April 1988. A total of 39 examinations were chosen, 20 from those which passed and 19 from those which failed. Each examination consisted of two writing tasks which had been scored holistically by a team of university raters, providing 78 separate writing samples to analyze for this study. In order to determine distinctions between the passing and failing groups, a two-level error analysis was conducted--one at the sentence level and another at the discourse (or rhetoric) level. The sentence error analysis

explored the errors in five broad categories: sentence, vocabulary, mechanics, verbs, and noun/pronoun. The five categories were further subdivided into a total of twenty-two features. The error frequency was recorded. The discourse analysis studied seven features: introduction, body, supporting details, connective elements, organization, and conclusion. On a scale of 0 to 2, the discourse features were rated according to an analytic scoring guide. The results of both analyses were calculated and correlated against the original holistic examination score and with each other. Ratios were determined on the frequency of normalized errors between the passing and failing groups, and correlation coefficients were calculated using the holistic exam scores. Ratios, normalized error frequencies and correlations were compared to the analyses scores to help interpret the significance of the errors to raters.

Results

Generally speaking, the results of the study show that the second language learners who failed the Oklahoma State University written proficiency examination did so for identifiable reasons when compared to the passing group: They didn't write as much; they wrote less well rhetorically; and they made nearly twice as many sentence errors as did the passing group. The number of words

written correlated moderately with the exam score (.60) and indicated that raters were influenced by apparent fluency. In addition, the study seemed to reveal feature distinctions which may be significant in determining differences between the passing and failing groups and/or important irritants to the raters. Features which seriously interfere with reader comprehension seemed to be significant sources of error to the second language learner. This study indicates that sentence errors in word choice concerning nouns and verbs and incoherent sentence construction errors seemed to irritate the rater to the point that they may make a difference between the passing and failing groups. Those who failed made at least twice as many errors in those features than those who passed, and those features correlated significantly with the exam score. Similarily, the discourse features which seemed to separate those who pass from those who fail are also central to reader rhetoric comprehension: body, organization, and connective elements. Those who passed scored significantly more points in these features than did those who failed. Finally, this study indicated a moderate correlation between the holistic examination score and the sentence error analysis (-.66) as well as between the exam score and the discourse analysis (.65). These correlations were possibly more significant than they seemed. The calculations done in this study were normalized for the number of words written, while the original holistic scores were not. Importantly, the two

analyses yielded almost the same correlations, indicating that raters gave no more weight to sentence errors than to rhetoric problems in general. Though they correlated only moderately well with each other, both analyses showed nearly the same percent/pass to percent/fail results as did the original holistic exam score.

Recommendations for Further Research

Further research needs to be done with university students who take English proficiency examinations. Studies undertaken to find distinguishing features determining pass or fail on writing done by native speakers might reveal those features which make non-native writing different from native student writing; or it could reveal further bias of the raters; or it might verify as universal problems those distinctive features found in the present study.

In addition, research on writing samples of non-native speakers who fell just below and just above the proficient/not proficient cut-off line might be useful to further narrow distinctive features which might cause one group to pass and one group to fail.

Further research needs to be done to verify the interrelationship of the three types of scoring methods used in this study. Perhaps the holistic scores on future writing proficiency examinations could be normalized and an absolute cut-off point be determined between passing and

failing writing samples. If this were done, correlations between separate sentence error analysis scores and discourse analysis scores such as done in this study could be compared to the normalized holistic examination scores to determine true significance.

Additional research with second language learners should involve those features which play a significant part in error gravity. If indeed raters seem to pay most attention to those features of words, sentences, and rhetoric patterns which interfere seriously with comprehension, then those features need to be identified and targeted for instruction.

Concerning pedagogical research, much is yet to be done to identify teaching methods which help the second language learner get at the meaning of words, sentences, and whole discourse. Further research is needed using experiments such as cognitive-based writing (Flower, 1979; Flower and Hayes, 1980, 1981), revision-based writing (Zamel, 1983, 1985), or process writing (Raimes, 1985, 1987; Celce-Muria, 1988). In addition, experiments could focus on the kinds of discourse schemas which would advance the second language learners' background knowledge--the kind of schemas which contain the words, sentence and rhetoric patterns needed to write with a native-like fluency.

REFERENCES

Applebee, A. N. (1981). Writing in the secondary school (NCTE Research Rep. No. 21). Urbanna, IL: National Council of Teachers of English.

Bhatia, A. T. (1974). An error analysis of students' compositions. International Review of Applied Linguistics, 12, 337-350.

Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). Research in Written Composition. Urbanna, IL: National Council of Teachers of English.

Burt, C.. (1917). The Distribution and Relations of Educational Abilities.

Burt, M. (1975). Error analysis in the adult EFL classroom. TESOL Quarterly, 9, 53-63.

Cast, B.M.D. (1939). The efficiency of different methods of marking English compositions. British Journal of Educational Psychology, 9, 257-269.

Celce-Muria, M. (1988). Integrating grammar into the ESL composition class. TESOL Newsletter, 6, 1-3.

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

Corder, S.P. (1967). The significance of learners' errors. International Review of Applied Linguistics, 5, 161-170.

Corder, S.P. (1974). Idiosyncratic dialects and error analysis. In J.C. Richards (Ed.), Error Analysis: Perspectives on Second Language Acquisition(pp. 158-171). London: Longman.

Coulter, K. (1968). Linguistic error-analysis of the spoken English of two native Russians. Unpublished masters thesis, University of Washington, Seattle.

Cummins, J. (1980). The construct of language proficiency in bilingual education. Current Issues in Bilingual Education. Washington, D. C.: Georgetown University Press.

Diederick, P., French, J., & Carlton, S. (1961). Factors in judgments of writing ability. Educational Testing Service Research Bulletin (R. B. 61-5). Princton, NJ: Educational Testing Service.

Flower, L. (1979). Writer-based prose: A cognitive basis for problems in writing. College English, 41, 19-39.

Flower, L. & Hayes, J. (1980). The cognition discovery: Defining a rhetorical problem. College Composition, 31, 21-32.

Flower, L. & Hayes, J. (1981). A cognitive process theory of writing. College Composition and Communication, 32, 34-324.

Harris, D. (1969). Testing English as a Second Language. New York: McGraw Hill.

Heaton, J. B. (1975). Writing English Language Tests. London: Longman Group.

Holderer, R. (1988). [Oklahoma State University writing proficiency examinations: Assessment scores]. Unpublished raw data.

Hunt, K. (1966) Cultural thought patterns in inter-cultural education. Language Learning, 16, 1-20.

Jain, M. P. (1974). Error analysis: Source, cause, and significance. In J. Richards (Ed.), Error Analysis: Perceptives on Second Language Acquisition (pp. 189-215). London: Longman Group.

Kaczmarek, C. M. (1980). Scoring and rating essay tasks. In J. W. Oller, Jr., and K. Perkins (Eds.), Research in Language Testing (pp. 151-159). Rowley, Mass: Newbury House.

Kaplan, R. B. (1987). Cultural thought patterns revisited. In U. Connor and R. Kaplan (Eds.), Writing Across Languages: Analysis of L2 Text. Reading: Addison-Wesley.

Lado, R. (1961). Language Testing: The Construction and Use of Foreign Language Tests. London: Longman.

McDaniels, B. A. (1985, March) Ratings vs equity in the evaluation of writing. Paper presented at the 36th annual meeting of the Conference on College Composition and Communication, Minneapolis, MN.

Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. TESOL Quarterly, 17, 651-671.

Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. TESOL Quarterly, 19, 229-258.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. Language Learning, 37, 439-468.

Richards, J. C. (1971). Error analysis and second language strategies. Language Sciences, 17, 12-22.

Richards, J. C., ed. (1974). Error Analysis: Perspectives on Second Language Acquistion. London: Longman.

Richards, J. C. (1974a). A non-contrastive approach to error analysis. In J. C. Richards (Ed.), Error Analysis: Perceptives on Second Language Acquisition (pp. 172-188). London: Longman.

Richards, J. C. (1974b). Social factors, interlanguage, and language learning. In J. C. Richards (Ed.), Error Analysis: Perceptives on Second Language Acquisition (pp. 64-91). London: Longman.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. TESOL Quarterly, 20, 83-93.

Rollins, M. F. (1985). Three factors in the written production of limited English writers in freshman composition courses at the University of Texas, El Paso. Masters thesis, University of Texas, El Paso.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics, 10, 219-231.

Shaughnessy, M. P. (1977). Errors and Expectations: A Guide for the Teacher of Basic Writing. New York: Oxford University Press.

Stiggins, R. J., & Bridgeford, N. J. (1983). An analysis of published tests of writing proficiency. Educational Measurement: Issues and Practices, 2, 6-19, 26.

Valentine, C. W. (1932). The Reliability of Examinations. London: Methuen.

Vann, R. J., Meyer, D. E., & Lorenz, F. O. (1984). Error gravity: A study of faculty opinion of ESL errors. TESOL Quarterly, 18, 427-440.

White, E. M. (1982). Comparison and contrast: The 1980 and 1987 California State University and Colleges Freshman Equivalency Examination. Long Beach: Trustees of the California State University.

White, E. M. (1984). Holisticism. College Composition and Communication, 35, 400-409.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. TESOL Quarterly, 17, 165-187.

Zamel, V. (1985). Responding to student writing. TESOL Quarterly, 19, 79-97.

Zughoul, M. R. & Kambal, M. O. (1983). Objective evaluation of EFL composition. International Review of Applied Linguistics in Language Teaching, 21, 87-103.

APPENDIXES

APPENDIX A

SAMPLE HOLISTIC SCORING GUIDE

Oklahoma State University
Scoring Instructions: English Proficiency Test
February 18, 1988

The test is in two parts which are to be scored separately. Part 1 will be scored on a scale of 6 to 1 (6 being the best), and Part 2 wil be scored on a scale of 4 to 1, (4 being the best). Each test will be scored by two readers, and the scores combined. When there is more than one point difference between the two readers' scores, a third reader will also score a paper and that score, if within one point of either of the first two, will count. When the third score is between two scores, such that either can be combined with it, the writer will receive the higher two scores.

A response receiving the best possible score on both parts from both readers will receive a total score of 20. A response receiving the lowest possible score on both parts from both readers will receive a total score of 4.

Instructions

1. Read the scoring guide and the assigned topics carefully.

2. Read all responses to Part 1 first, scoring as you read, and recording your scores, along with the booklet number, on the slips of paper provided.

3. As always, students should be rewarded for what they do well in response to the topic. Remember that you are reading a rough draft prepared under serious time constraints. Responses that argue with the assigned topic or treat it superficially should customarily receive a score no higher than 2, regardless of how well written.

4. Remember that the exemplary response may take an unusual or creative approach to the topic.

SCORING GUIDE

Part 1

This assignment asks students to describe a change that they would make if they were able (if they had the authority or power to do so) and argue convincingly that its effects would be an improvement over condiditons that exist now. (Students were instructed that they could write about any change in any setting they chose.)

6 A response that receives this score will respond to all parts of the assignment, describing a change and arguing that its effects will be an improvement over conditions that

exist now. This implies that the student will describe both the effects and the existing conditions to some extent. It will present pertinent, detailed examples and will be orderly, moving smoothly from point to point. It will use language in an exemplary way, displaying few if any problems with syntax, usage, or textual conventions.

5 A response that receives this score will respond to all parts of the assignment, but less thoroughly or less well than a response that receives a 6. For example, the description of the change or the argument that its effects will be an improvement over existing conditions may be less clear or less detailed than that found in a paper that receives a 6. The response will be well organized, and although the writer may have some problems with syntax, usage, or textual conventions, these will not be serious lapses.

4 A response that receives this score may slight but not ignore one or more of the assigned tasks, or treat all tasks less thoroughly than a response that receives a 5. For example, its description of the change may be incomplete, or the entire response may be less detailed than one that receives a score of 5 or 6. The argument that the change will result in an improvement may be less detailed or less convincing, and its examples and illustrations may be less detailed than those in a response that receives a 5. Its use of language (syntax, usage, and textual conventions) will be less skilled than a 5.

3 A response that receives this score may slight or ignore one or more parts of the assignment; for example, it may select a change but not describe it in adequate detail, or its argument may be less than convincing. It may give few examples or illustrations, and these not richly detailed. It may show that the writer has difficulty with organization, and its handling of syntax, usage, and textual conventions will show some serious problems, but not to the extent that they interfere with readability.

2 A response that receives this score will ignore more than one part of the assignment. Its description of a change may be incomplete or superficial, it amy provide few or no examples and illustraions, or those present may be incomplete or lack sufficient detail. It may ignore the instruction to argue that the change will improve existing conditions. Such a response will show serious problems with organization, and its control of usage, suntax, and textual conventins will show serious difficulty throughout.

1 A paper that receives this score will not really engage the topic. It will show almost no organization, and its control of language and textual conventions will create the impression on ineptitude.

PART 2

This assignment asks students to repsond to the following topic: People often say that if they could do something over again, they would do it differently. Describe a choice or decision you have made that you would not make differently, and explain why you would act differently now than you did before.

Successfully completing this assignment requires students to shift from description to exlanation, and this move up at least one level of abstraction should be a key distinction between upper half and lower half responses. Responses that argue with the assigned topic or treat it superficially should not normally receive a score higher than a 1, regardless of how well written.

Criteria for Scores

4 A response that receives this score will complete both parts of the assignment, describing a choice or decision the writer has made and explaining why the writer would act differently if faced with that same choice or decision again. The description will be supported by details, examples, and illustrations. Such a response will be orderly and move smoothly from point to point. It will display exemplary control of language and textual conventions.

3 A response that receives this score will differ from a 4 in the completeness with which it describes the choice or decision or in the number and quality of examples and details that illustrate the choice or the change that the writer would make. The response may slight, but not ignore, one part of the assignment. It will be organized, but not as effecively as a 4, and its control of language may show lapses in control of syntax, usage, or textual conventions. These will not be serious lapses, nor will they interfere with readability.

2 A response that receives this score may ignore one part of the assignment, and it may slight the other, perhaps describing the choice or explaining the change but the writer would make in less detail than a response that receives a 3. It may show major difficulties with organization or control of language.

1 A response that receives this score may treat the assignment superficially, or may argue with the topic. It may be poorly organized, if at all, and its control of usage, syntax, and textual conventions may create the impression of ineptitude.

APPENDIX B

COMPUTER PRINTOUT OF RAW DATA

ERROR AND DISCOURSE ANALYSIS SUMMARY

RESULTS- SUMMARY

Stu. No.	NoWords	TotErr	N-e	DisSc	ExScore	EssayNo
P1	701	91	130	16	8	2
P2	1124	97	86	18	17	3
P3	1007	161	160	11	10	4
P4	1051	116	110	17	12	4
P5	599	17	28	14	11	5
P6	1086	62	57	21	11	6
P7	665	108	162	11	11	6
P8	754	149	198	17	9	9
P9	963	17	18	17	16	9
P10	778	97	125	15	9	9
P11	1369	127	93	19	16	9
P12	961	52	54	23	14	9
P13	1015	85	84	13	13	9
P14	782	70	90	15	9	10
P15	557	37	66	7	9	10
P16	963	135	140	16	9	10
P17	741	123	166	22	9	10
P18	1230	160	130	16	9	10
P19	1355	92	68	22	14	10
P20	955	62	65	17	14	10
F1	1070	147	137	9	9	1
F2	763	183	240	13	7	1
F3	810	134	165	12	7	4
F4	817	128	157	10	7	6
F5	805	238	296	11	5	6
F6	560	76	136	9	10	7
F7	705	301	427	8	6	8
F8	1006	173	172	10	6	8
F9	611	107	175	9	5	9
F10	848	207	244	6	8	9
F11	463	117	253	6	6	10
F12	670	111	166	13	8	10
F13	627	77	123	8	6	10
F14	599	60	100	13	9	11
F15	674	77	114	17	10	11
F16	492	116	236	4	7	11
F17	810	50	62	14	9	12
F18	857	101	118	15	9	12
F19	696	99	142	8	7	12
TotErr	32539	4360	5491.85	522	371	
Mean	834.33	111.79	140.82	13.38	9.51	
Range	906	284	409.30	19	12	
StdDev	220.37	56.62	77.99	4273	3.03	
rScore	0.60	-0.47	-0.66	0.65	0.00	

ERROR ANALYSIS BY CATEGORY

RESULTS- ERROR ANALYSIS

Stu. No.	Vocab	N/PProb	Verb	Sent	Mech	ExScore	EssayNo
P1	44	8	8	9	22	8	2
P2	37	6	9	8	37	17	3
P3	74	26	6	12	43	10	4
P4	67	20	6	4	19	12	4
P5	7	3	3	0	4	11	5
P6	25	1	5	2	29	11	6
P7	23	0	0	6	79	11	6
P8	71	32	12	6	28	9	9
P9	7	1	5	0	4	16	9
P10	49	15	7	6	20	9	9
P11	34	16	26	8	43	16	9
P12	11	10	6	4	21	14	9
P13	24	10	7	4	40	13	9
P14	28	10	14	2	16	9	10
P15	16	6	8	3	4	9	10
P16	55	17	9	20	34	9	10
P17	53	13	22	9	26	9	10
P18	66	20	3	18	53	9	10
P19	36	32	31	22	24	14	10
P20	9	8	0	0	45	14	10
F1	37	20	8	7	75	9	1
F2	86	33	17	17	30	7	1
F3	45	22	20	12	35	7	4
F4	50	19	20	11	28	7	6
F5	107	22	21	22	66	5	6
F6	18	1	5	2	50	10	7
F7	73	15	16	21	176	6	8
F8	93	20	16	18	26	6	8
F9	48	20	9	11	19	5	9
F10	97	47	20	10	33	8	9
F11	28	11	8	8	62	6	10
F12	45	19	22	3	22	8	10
F13	42	2	11	10	12	6	10
F14	36	2	1	7	14	9	11
F15	32	13	13	3	16	10	11
F16	25	3	1	4	83	7	11
F17	19	9	10	1	11	9	12
F18	37	16	6	11	31	9	12
F19	53	15	4	12	15	7	12
TotErr	1707	563	415	333	1395	371	
Mean	43.77	14.44	10.64	8.54	35.77	9.51	
Range	100	47	31	22	172	12	
StdDev	25.07	10.29	7.53	6.33	30.16	3.03	
N-e	52.46	17.30	12.75	10.23	42.87		
%TotErr	39.2	12.9	9.5	7.6	32.0		
rScore	-0.51	-0.24	-0.11	-0.40	-0.20	1.00	

CATEGORY: VOCABULARY

RESULTS- ERROR ANALYSIS

Stu. No.	WC Prep	WC Conn	WC Art	WC Mod	WC Aux	WC N/V	W Form	WF-ing-ed
P1	6	5	16	2	0	5	4	6
P2	3	1	10	1	1	5	6	10
P3	8	10	21	3	15	10	0	7
P4	9	1	24	5	8	9	4	7
P5	4	0	1	0	0	1	1	0
P6	3	1	18	1	0	1	0	1
P7	3	4	7	1	0	6	1	1
P8	14	4	37	3	3	4	2	4
P9	1	0	2	1	1	0	1	1
P10	5	1	17	1	8	8	2	7
P11	6	3	10	4	1	4	0	6
P12	0	0	3	0	0	2	0	6
P13	5	0	9	0	1	6	1	2
P14	4	4	8	2	2	0	5	3
P15	1	0	4	0	5	1	0	5
P16	7	2	20	4	4	12	0	6
P17	12	6	8	1	4	6	4	12
P18	18	5	8	1	1	20	6	7
P19	4	1	9	1	5	9	0	7
P20	1	0	1	0	5	0	2	0
F1	3	1	10	7	0	7	2	7
F2	9	0	26	2	6	19	12	12
F3	7	1	13	3	5	11	0	5
F4	8	1	16	2	3	7	8	5
F5	18	6	22	6	9	21	17	8
F6	6	0	5	0	0	5	0	2
F7	7	9	21	7	2	16	6	5
F8	19	2	42	3	0	13	11	3
F9	6	3	9	8	4	8	5	5
F10	14	2	31	5	3	26	6	10
F11	8	3	9	0	2	4	1	1
F12	7	1	15	2	3	9	3	5
F13	9	4	6	6	1	11	1	4
F14	5	3	13	2	1	8	2	2
F15	9	1	7	4	2	3	1	5
F16	2	2	9	1	2	7	2	0
F17	2	0	5	0	1	3	2	6
F18	7	4	4	3	4	5	2	8
F19	1	0	9	4	3	14	8	14
TotErr	261	91	505	96	115	306	128	205
Mean	6.69	2.33	12.95	2.46	2.95	7.85	3.28	5.26
Range	19	10	41	8	15	26	17	14
StdDev	4.73	2.44	9.43	2.21	3.07	6.09	3.76	3.39
N-e	8.0	2.8	15.5	3.0	3.5	9.4	3.9	6.3
%TotErr	5.99	2.09	11.58	2.20	2.64	7.02	2.94	4.70
rScore	-0.45	-0.33	-0.36	-0.43	-0.15	-0.49	-0.44	-0.12

CATEGORIES: N/P VERB

RESULTS- ERROR ANALYSIS

Stu. No.	UnclrPro	Agr Pro	Poss	Number	Agr S/V	Tense
P1	2	0	0	6	5	3
P2	1	0	1	4	4	5
P3	3	0	2	21	5	1
P4	4	0	0	16	4	2
P5	0	0	0	3	0	3
P6	0	0	0	1	1	4
P7	0	0	0	0	0	0
P8	1	0	0	31	2	10
P9	0	0	0	1	5	0
P10	0	2	1	12	2	5
P11	0	3	0	13	16	10
P12	0	0	2	8	2	4
P13	2	1	3	4	1	6
P14	0	0	0	10	2	12
P15	0	0	0	6	2	6
P16	3	0	2	12	2	7
P17	1	0	0	12	4	18
P18	1	0	4	15	2	1
P19	0	0	0	3	1	9
P20	0	0	5	3	0	0
F1	1	1	0	18	5	3
F2	2	2	2	27	13	4
F3	2	0	0	20	3	17
F4	6	2	0	11	6	14
F5	3	1	0	18	5	16
F6	1	0	0	0	3	2
F7	8	0	0	7	3	13
F8	1	0	2	17	11	5
F9	16	1	0	3	4	5
F10	6	2	3	36	14	6
F11	0	0	0	11	3	5
F12	0	3	0	16	9	13
F13	0	0	0	2	6	5
F14	0	0	0	2	0	1
F15	4	1	1	7	5	8
F16	2	0	0	1	1	0
F17	0	2	0	7	8	2
F18	3	1	2	10	3	3
F19	1	0	1	13	2	2
TotErr	74	22	31	407	164	230
Mean	1.90	0.56	0.79	10.44	4.21	5.90
Range	16	3	5	36	16	18
StdDev	2.99	0.90	1.26	8.51	3.80	4.92
N-e	2.3	0.7	1.0	12.5	5.0	7.1
%TotErr	1.70	0.50	0.71	9.33	3.76	5.28
rScore	-0.39	-0.06	0.18	-0.30	-0.09	-0.26

CATEGORY: SENTENCE MECHANICAL

RESULTS- ERROR ANALYSIS

Stu. No.	Parallel	IncConst	Frag	RunOn	Spelling	Punct	Comma	Cap
P1	0	9	3	1	14	1	1	2
P2	0	8	2	1	9	4	12	9
P3	0	12	3	8	17	1	13	1
P4	0	4	1	0	7	1	6	4
P5	0	0	0	0	1	1	2	0
P6	0	2	0	4	6	3	14	2
P7	0	6	4	29	21	2	18	5
P8	0	6	2	4	8	1	5	8
P9	0	0	0	0	1	0	3	0
P10	0	6	2	5	2	8	3	0
P11	1	7	1	8	16	5	13	0
P12	1	3	1	1	6	0	10	3
P13	0	4	2	2	11	7	12	6
P14	0	2	2	2	12	0	0	0
P15	0	3	0	0	4	0	0	0
P16	1	19	3	1	19	2	7	2
P17	3	6	5	4	7	1	3	6
P18	2	16	5	2	9	14	6	17
P19	3	2	3	4	6	2	5	18
P20	0	0	0	1	16	4	23	1
F1	0	7	1	6	30	12	12	14
F2	3	14	5	7	6	1	6	5
F3	2	10	5	19	2	1	8	0
F4	0	11	2	4	14	1	6	1
F5	0	22	11	6	16	0	15	18
F6	0	2	3	1	27	6	7	6
F7	0	21	8	21	48	7	87	5
F8	0	18	4	8	6	4	1	3
F9	0	11	3	3	2	3	7	1
F10	2	8	10	3	11	1	4	4
F11	0	8	4	5	31	2	18	2
F12	0	3	7	1	5	0	1	8
F13	0	10	3	2	2	0	5	0
F14	0	7	0	2	5	1	6	0
F15	0	3	0	2	8	0	5	1
F16	0	4	1	5	21	1	7	48
F17	0	1	0	1	7	0	3	0
F18	0	11	2	7	11	1	10	0
F19	0	12	1	4	5	3	2	0
TotErr	18	298	109	184	449	101	366	200
Mean	0.46	7.64	2.79	4.72	11.51	2.59	9.38	5.13
Range	3	22	11	29	47	14	87	48
StdDev	0.93	5.78	2.65	5.90	9.67	3.22	13.69	8.58
N-e	0.6	9.2	3.3	5.7	13.8	3.1	11.2	6.1
%TotErr	0.41	6.83	2.50	4.22	10.30	2.32	8.39	4.59
rScore	0.05	-0.56	-0.48	-0.20	-0.16	0.07	-0.07	-0.09

DISCOURSE ANALYSIS SUMMARY

RESULTS- DISCOURSE ANALYSIS

Stu.No.	Intro	Th.State	Body	Details	Conn.	Organ.	Concl	Total	ExScore	EssayNo
P1	1	2	3	2	4	4	0	16	8	2
P2	4	2	3	2	4	3	0	18	17	3
P3	1	2	1	2	1	2	2	11	10	4
P4	1	4	4	1	1	4	2	17	12	4
P5	3	3	0	2	3	2	1	14	11	5
P6	2	4	4	3	2	3	3	21	11	6
P7	1	2	3	1	1	3	0	11	11	6
P8	1	2	2	1	4	4	3	17	9	7
P9	0	3	3	2	3	4	2	17	16	7
P10	2	1	3	1	2	4	2	15	9	7
P11	2	3	3	2	3	4	2	19	16	7
P12	1	3	4	3	4	4	4	23	14	7
P13	2	0	2	2	2	1	4	13	13	7
P14	1	3	3	3	2	3	0	15	9	10
P15	1	2	2	0	1	0	1	7	9	10
P16	2	1	3	3	2	2	3	16	9	10
P17	4	4	2	3	3	3	3	22	9	10
P18	2	2	4	1	3	2	2	16	9	10
P19	2	3	4	3	3	4	3	22	14	10
P20	0	2	4	2	4	4	1	17	14	10
F1	1	2	2	2	1	1	0	9	9	1
F2	1	2	1	2	3	2	2	13	7	1
F3	1	2	2	2	1	3	1	12	7	4
F4	2	1	1	2	1	2	1	10	7	6
F5	0	1	2	3	2	2	1	11	5	6
F6	2	3	0	1	0	1	2	9	10	7
F7	1	0	2	2	1	2	0	8	6	8
F8	0	2	0	2	2	3	1	10	6	8
F9	1	1	2	0	2	2	1	9	5	9
F10	1	1	0	2	1	0	1	6	8	9
F11	0	2	2	0	0	1	1	6	6	10
F12	1	2	3	1	2	2	2	13	8	10
F13	3	2	1	1	1	0	0	8	6	10
F14	3	2	2	2	2	2	0	13	9	11
F15	4	3	1	3	2	2	2	17	10	11
F16	1	1	0	1	0	0	1	4	7	11
F17	2	2	2	1	3	2	2	14	9	12
F18	2	1	2	2	2	3	3	15	9	12
F19	1	1	1	0	2	1	2	8	7	12
TOTALS	60.00	79.00	83.00	68.00	80.00	91.00	61.00	522.00	371.00	
MEAN	1.54	2.03	2.13	1.74	2.05	2.33	1.56	13.38	9.51	
STD.DEV	1.06	0.97	1.22	0.90	1.13	1.25	1.13	4.73	3.03	
r SCORE	0.23	0.44	0.50	0.30	0.49	0.50	0.31	0.65	1.00	

VITA

Jeana M. S. Johnson

Candidate for the Degree of

Master of Arts

Thesis: AN ANALYSIS OF THE PERFORMANCE OF NON-NATIVE UNIVERSITY STUDENTS ON AN ENGLISH PROFICIENCY EXAMINATION

Major Field: English

Biographical:

Personal Data: Born in Long Beach, California, December 31, 1939, the daughter of Mr. and Mrs. William Sanford Snodderly. Married David E. Johnson, August 26, 1961; one daughter, Kathleen.

Education: Graduated from Fulton High School, Knoxville, Tennessee, 1957; received Bachelor of Arts Degree in English Education from Louisiana State University at New Orleans in 1971; National Endowment for the Humanities graduate scholarship recipient in Spanish, summer 1986, Oklahoma State University; completed requirements for the Master of Arts Degree at Oklahoma State University in May 1989.

Professional Experience: English as a Second Language teacher, San Tome kindergarten, San Tome, Venezuela, 1971-73, 1975; English teacher, York Academy, Shacklesford, VA, 1974-75; English teacher, Woodward Junior High, Woodward, OK, 1977-80; English teacher, Tulsa Junior College, 1980-81; English/Spanish teacher, Liberty High School, Tulsa, OK, 1982-86; Teaching Assistant, English Department, Oklahoma State University, 1986-87; Teaching Assistant for English as a Second Language, English Department, Oklahoma State University, Spring, 1988; Member NCTE, OCTE, OKTESOL, TESOL.