The effect of error on the grading of ESL and native-speaker freshman writing: A comparison

Rita Kathlyn Somers

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd-project

Part of the First and Second Language Acquisition Commons

Recommended Citation

https://scholarworks.lib.csusb.edu/etd-project/812

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
THE EFFECT OF ERROR ON THE GRADING OF ESL AND NATIVE-SPEAKER FRESHMAN WRITING: A COMPARISON

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
English Composition

by
Rita Kathlyn Somers

December 1993
THE EFFECT OF ERROR ON THE GRADING OF ESL AND NATIVE-SPEAKER FRESHMAN WRITING: A COMPARISON

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

by
Rita Kathlyn Somers
December 1993

Approved by:

Wendy Smith, Chair, English

Dian Pizyrie

Margaret Doane

April 21, 1994
Dedicated with much love and gratitude to all of my family, whose concern and support have greatly assisted in the completion of this project.
ACKNOWLEDGEMENTS

I wish to express my most sincere thanks and gratitude to several people who have encouraged and assisted me during this project. First, to my dearest friend, Wilt Senecal, I am particularly thankful for his encouragement, his support, and for always being there during this endeavor. Also, I am most appreciative to my family, particularly my parents, for their unfailing love, concern, and assistance during the course of this project, and for never letting me forget how proud they are of me.

Next, I would like to thank Professor Wendy Smith for suggesting this research topic and for advising me on questions regarding methodology. Also, I thank her for serving as my chief committee reader, intern advisor, and instructor. I also thank the other members of my committee for their guidance and support.

This project would not have been possible without the approval and assistance of the freshman common reading/writing committee. I wish to express appreciation to the committee for allowing me to collect data from the classes and to its chair, Loralee McPike, for her assistance in recruiting raters for my project. Thanks also go to instructors Wendy Smith and Phyllis Gilbert for permitting their students to participate in this study.
In addition, for scoring the essays involved in this project, I would like to thank the following instructors: Professors Loralee McPike, Cynthia Pastrana, Phil Page, and Milton Clark.

Also, special thanks go to Richard Newman of the University of California at Riverside for his assistance and consultation regarding the statistical analyses used in this study. I am also grateful to Professor Rong Chen of California State University, San Bernardino for assistance regarding statistical analyses.

A special note of thanks goes to Robert Land, Jr. of the University of California at Irvine for conversations involving the findings of similar projects.

Finally, I would like to thank Alex McKeown-Boyd and Berne Mulvihill of the CSUSB Learning Center's tutoring program for their assistance.
# LIST OF TABLES AND FIGURES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Descriptive Statistics (Uncorrected)</td>
</tr>
<tr>
<td>2</td>
<td>Frequency Distribution of Scores (Uncorrected)</td>
</tr>
<tr>
<td>3</td>
<td>Descriptive Statistics (Corrected)</td>
</tr>
<tr>
<td>4</td>
<td>Frequency Distribution of Scores (Corrected)</td>
</tr>
<tr>
<td>5</td>
<td>MANOVA Tables</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Graph of Uncorrected Scores</td>
</tr>
<tr>
<td>2</td>
<td>Graph of Corrected Scores</td>
</tr>
<tr>
<td>3</td>
<td>Mean Score of the Four Groups</td>
</tr>
</tbody>
</table>
ABSTRACT

Comparisons between English and ESL texts could help determine whether essays from comparable courses are rated similarly. This study examines whether surface errors affect ESL holistic scoring. It compares CSUSB freshman composition essays based on the same topic written by native English speakers and ESL students.

Essays from the same term’s common reading/writing assignment, written under similar conditions, were compared for morpho-syntactic errors. Both sets of papers were holistically scored by trained English department raters, with third raters used for score deviations of more than two points. After surface errors were corrected, the texts were scored again, by other raters. The four sets of data were then examined for variations in scoring between the two groups, based on surface errors and on scores for before and after correction.

The results showed corrected scores for both groups significantly higher than prior scores. There were significantly lower scores for ESL papers as compared to native speaker papers before corrections were made; however, there was no significant difference between the two groups after correction. This indicates that raters do not always downgrade ESL essays for features other than surface errors.
CHAPTER ONE

INTRODUCTION

Much has been researched and discussed about instructors' responses to the writing of English as a Second Language (ESL) students. Researchers report findings on studies which examine teacher comments on essays, attitudes toward early drafts of papers, and opinions toward error. Moreover, some studies report on response differences between native-speaker (NS) and non-native-speaker (NNS) student writing. Other studies involve teacher responses to morpho-syntactic and mechanical errors, while still others deal with their response to such composing issues as content and rhetorical matters.

Zamel (1985), for one, states that teacher response to NNS compositions focuses primarily on surface-[sentence] level features of writing rather than on other writing issues. However, according to Green and Hecht (1985), readers cannot agree on the establishment of a corpus of errors or their respective gravity because they subscribe to differing models of correctness. This lack of agreement can cause inconsistencies in ESL holistic evaluations.

However, others feel that widely divergent scores among ESL papers are caused by factors other than surface errors. Ruetten (1991) states that problematic, holistically-scored
ESL papers (i.e., essays that cause great variation in ratings among readers) often do not meet reader expectations. In other words, differences in some facets of ESL writing such as content and syntax influence some raters more than others. For this reason, some teachers respond negatively in their scoring while others do not. Freedman (1979a), in turn, states that cultural experience is a factor in the development of a writer's topic. This variation in cultural experience causes ESL essays to seem "less academic" to the U.S. discourse community, thus lowering their scores.

Researchers also report various findings concerning reader response to writing by different categories of readers. By way of illustration, both James (1977) and Sheorey (1986) discovered different hierarchies of error gravity in the responses of NS and NNS instructors of English. According to a study by Siegel (1982), teachers new to the field of English often ignore some more serious errors and mark five times as many unnecessary or mistaken surface "corrections" as do experienced English teachers.

Are essays from students in comparable-level English and ESL courses, then, graded similarly? Little attention has been directed toward reader-response comparisons of ESL texts with NS texts.
OBJECTIVES OF THE STUDY

The focus of the research and study for this thesis is on determining the role of error in teacher evaluation of essays. In particular, this study examines comparable writing samples of NS’s and NNS’s. I have applied to the two groups of writing samples knowledge gained from error analysis in written composition and statistical methods of comparison.

I hope that this study, although small and limited in scope, will add to the slight body of research to date concerning the role of error in instructor response to the writing produced by ESL students, particularly in comparison with their NS classmates.

The study’s objectives are to examine and compare the current NS and NNS research regarding error, rhetorical methods, and reader response in order to determine findings regarding any notable differences in writing, holistic score comparisons, and/or coherence issues. The purpose of this study, then, is to examine whether raters holistically score the writing of ESL students lower than that of NS’s writing on the same prompt and, if they do, whether this is due to surface errors unique to ESL students or to larger rhetorical issues.

This study stems from the research done by Fein (1980), McGirt (1984), Whitley (1984), and Sweedler-Brown (1993a,
1993b). The study examines their findings regarding NS's and NNS's related to notable differences in writing (Fein 1980), ratings for before and after correction of ESL papers (Sweedler-Brown 1993a), and holistic score comparisons between the two groups of writers before and after correction of surface errors (McGirt 1984, Whitley 1984, Sweedler-Brown 1993b).

**RESEARCH QUESTIONS**

The following questions comprise the focus of the research for this study: (1) Is there more variation in holistic scores awarded to ESL essays than to NS essays, in other words, less interrater reliability?; (2) With surface errors corrected, do the ratings increase comparably for both groups?; and (3) Do raters respond differently to surface errors more or to some other non-native quality in the writing, as judged by their response to corrected, error-free compositions?

The current study compares the scores of two groups of papers (N=34) written during the winter, 1992 term for the California State University, San Bernardino (CSUSB) mid-quarter, common reading/writing assignment for two comparable freshman English classes, English 101 and ESL English 101. Each paper was holistically scored. Then, after morpho-syntactic (surface) error correction, the texts were
scored holistically again by two other raters. Next, I examined the data for noticeable variations between scores for corrected and uncorrected papers between the two groups.

My hypothesis was that the NNS scores would be lower than the NS scores before correction, but that the NNS ratings would exhibit a corrected score comparatively higher than that for NS scores after correction. However, I still expected the corrected NNS ratings to remain significantly lower than the NS ratings. This hypothesis was made based on the majority of available findings (Fein 1980; McGirt 1984; Whitley 1984; Sweedler-Brown 1993a; Land & Whitley 1989, p. 286), which state that readers do indeed penalize NNS writing for surface errors despite the fact that those errors do not obscure the writer’s message.

Another reason for the hypothesis was that the majority of studies done thus far on rhetorical issues and the few done on comparisons of NS and NNS writing indicate that NNS students score lower than NS even on error-free papers (Fein 1980; McGirt 1984; Whitley 1984; Sweedler-Brown 1993a, 1993b; Santos 1988; Ludwig 1982).

OUTLINE OF THE STUDY

The prior research that tests this hypothesis and responds to the three research questions is set forth next in Chapter II, which presents a review of the literature.
Issues examined include reader responses to writing and the treatment of error. Also, I investigate factors other than surface errors which affect readers, and I examine hierarchies of error gravity.

Next, methods of evaluating writing, particularly holistic scoring, are discussed. In conclusion is a comparison of NS and NNS texts. This section details the explorations of Fein (1980), McGirt (1984), and Sweedler-Brown (1993a), whose inquiries parallel this present study, and of another study done by Sweedler-Brown (1993b), in which she compared NNS scores before and after correction (without comparisons with NS scores).

Then, Chapter III presents a background of the study, noting the CSUSB context, a description of the two comparable freshman composition courses, and placement of NNS students in either English 101 or ESL English 101. A description of both groups of students is given, as is the selection of the data base. Procedures for identifying, analyzing, correcting, and reporting the results conclude this chapter.

Results of the study are presented in Chapter IV. I examine and compare uncorrected and corrected holistic scores, and I discuss the need for third raters. Next, comparisons of the four groups of papers are made. The findings, including those concerning interrater reliability,
are discussed. Then I examine the results for each of the three main research questions.

In conclusion, Chapter V discusses the results of the study. This chapter deals with the writing proficiency of both NS and NNS students. This final chapter follows with a discussion of the effect of the writing of both groups of students upon reader/raters. Then, Chapter V offers a summary of the findings. Finally, the chapter concludes with theoretical and pedagogical implications and with suggestions for further research.
CHAPTER TWO

REVIEW OF THE LITERATURE

INSTRUCTOR RESPONSE TO WRITING

Much research exists concerning how teachers respond to the writing efforts of their students. Some studies discuss scores they assign, while others deal with instructor comments. Zamel (1985) states that teachers note similar comments on many of their students' papers and apply a single ideal writing standard to all papers. Their comments, she notes, are vague and abstract. In addition, readers of student compositions will find errors if that is what is sought during a reading (Sommers 1982). Sommers states:

For the most part, teachers do not respond to student writing with the kind of thoughtful commentary which will help students to engage with the issues they are writing about or which will help them think about their purposes and goals in writing a specific text. (Sommers, 1982, p. 154)

Additionally, Sommers calls for teachers to reconsider their purpose for, and thus their response to, a first draft of a paper. Most instructors respond to such papers as if they were final drafts rather than a tentative beginning or an exploration of a topic.

Furthermore, teachers would be more helpful if they were more specific in their comments to students about what is wrong with a paper (Sommers 1982). According to Zamel
(1985), instructor responses are not only vague but are also inconsistent in reference to error.

Rather than reading for errors, teachers might engage with a student's paper for one of four other reasons: to read and respond, judge, analyze the work in the manner of a literary critic, or assist the student to improve as a writer rather than to improve the text itself (Purves 1984). Ideally, the role of the instructor as a reader offers four choices for response. These options are to focus on the content, organization and presentation, style and tone, or to express the teacher's interest or personal response to the text. The choice of one of these responses should be guided by the kind of writing assigned and the cultural background of the student, according to Purves (1984).

Another researcher's method of response to student writing is that reported on by Hake (1986). She finds that raters of written examinations respond differently to pure narration than they do to narration set within exposition (p. 160). This finding indicates that instructors need to determine what they are testing for, evidence of the rhetorical skill of the writer or evidence of the writer's ability to meet the framework criteria of the rater. Student writers, then, should be informed not only of the difference between the two methods of writing but also of the preference of the rater for each assignment.
RESPONSE TO ESL WRITING

Another important issue to examine relative to teacher response to writing is that toward the compositions of NNS students. In particular, often NNS readers respond differently to student papers than do NS readers. Many researchers (Sheorey 1986, Santos 1988, James 1977) have found that NS raters grade ESL papers less severely than do NNS raters.

However, results in this area are often conflicting. For one example, other work by Land and Whitley (1989, p. 287) reports the opposite: that NS's grade more severely. Moreover, in research done by Ludwig (1982), which involved having teachers watch a student give a presentation on a videotape, NS raters not only graded the NNS's more harshly, but also began writing down grades before the student whom they were watching had finished his or her presentation.

While NS's write more messages to students on papers than do NNS's, both groups often neglect the communicative dimension of evaluation (Green & Hecht 1985). Moreover, non-teacher readers of student discourse react similarly. In evaluating the English papers of Greek students (Hughes & Lascaratou 1982), both NS and NNS raters, including non-teachers, awarded higher ratings than did NS Greek teachers. The English NS group of raters seemed to stress student intelligibility more than correctness. Of all groups of
raters, though, non-teacher NS’s seemed to be the ones who are most accepting of second language written communications (Ludwig 1982).

Of all groups of readers who are teachers, studies indicate factors other than native language may influence how a teacher responds to ESL papers. For example, Vann et al (1984) find that the age and academic discipline of a teacher will influence the response. It has also been discovered that instructors from departments other than English respond with a wide variation (Siegel 1982). Furthermore, one report states that younger teachers and those who have undergone less rigorous academic programs grade less severely than other teachers (Ludwig 1982). One study even found that female raters appear to be more conservative than their male colleagues in scoring (Hairston 1981). Thus, a wide range of types of reader response exists.

Some researchers have explored teacher attitudes toward the perceived purpose of ESL student writing. In other words, these studies examine teachers’ responses to the content, or meaning, of ESL papers rather than to the form in which the compositions are fashioned.

In India, a study of one component of the Bangalore/Madras Communicational Teaching Project (which stresses unconscious processes), which examined a current
model of language learning, reports that many more content than form errors were attended to by instructors (Beretta 1989). Since teachers cannot know the cognitive processes of students, the theory behind the project states, they cannot preplan lessons. Another study which examined teacher response (Searle & Dillon 1980), however, reveals that teachers respond overwhelmingly to form rather than content. This study found, in the comments made by readers on papers, few attempts to encourage the development in students of thinking through their writing.

Fathman and Whalley (1990) also investigated the value of feedback in the writing done in intermediate ESL composition classes. Four methods of instructor feedback were offered to portions of the population: grammar, content, grammar and content, and no feedback. The students in all four groups improved both grammar and content after rewrites in response to any feedback received. The researchers found, however, that grammar feedback does not affect the content of NNS papers to any appreciable extent (p. 183). In this manner, these researchers conclude that the order of grammar or content instruction does not affect student writing ability, nor does simultaneous teaching of the two categories confuse ESL students (p. 185). The act of rewriting, then, is a helpful practice for improving writing, regardless of instruction received.
Another study (Freedman 1979a) examined what component of papers is most important to raters. By rewriting the content, organization, sentence structure, and mechanics of papers in different versions for raters, Freedman found that teachers rate content and organization as the most important component. Nevertheless, in their comments to students, the raters most frequently mentioned mechanics. In addition, the teachers sometimes wrote incorrect changes in their comments to students (Freedman 1979b).

Santos (1988), however, in her study found the content of writing rated lower than the form. Mullen (1980, p. 160), in her study of teacher ratings, discovered another variation: that vocabulary usage was most important and organization the least important. Thus, we see that a wide range of reader responses to ESL writing exists.

CORRECTNESS

To some readers, correctness is the hallmark of acceptable writing, and perfection in the usage of standard English defines correctness. In a further examination of who tolerates which variations in writing, Hairston (1981) found that readers who are in the professions (mostly business executives and attorneys) do care about at least some parts of standard English usage. Moreover, Shuy (1976, p. 313) declares that mainstream American society tolerates
phonological much more than grammatical variation. In the academic setting, the required mode of writing has long been the grammar of standard written English.

At the same time, attempts to obtain agreement on a definition of correctness are generally unsuccessful. Carlson and Bridgeman (1986) affirm that no single measure of correctness in writing exists and that any attempts to assess it have been restricted to judgments regarding grammaticality (p. 129). Does correctness mean error-free writing? To many instructors, such is the definition of correctness, for the avoidance of error does predict the ability to manage successfully complex sentence structures (Carlson & Bridgeman 1986, p. 144).

Correctness, then, often is equated less with comprehensibility and more with grammatical perfection of writing; i.e., with a lack of error, according to many instructors. Unfortunately, this view of correctness often is applied by teachers to the evaluation of their students' writing despite current findings about writing to the contrary. In the minds of some instructors, the "content vs. process" or "form vs. meaning" debate still has not been settled in favor of the message and the thought that has produced it.

In an examination of the purpose of writing, either to present a sober, well-thought out and well-argued paper (an
expressivist axiology) or an error-free paper (a formalist axiology), little agreement exists regarding how teachers should react to what their students have produced and on what their feedback should focus (Fathman & Whalley 1990, p. 178). According to Chastain (1981), the purpose of writing is to serve as communication. For many teachers, writing should serve as a means of discovery of the writer's viewpoint, achieved through a multi-step process of various drafts of a paper. This outlook, in other words, promotes both self awareness and communication.

ERROR

Surface Errors

The two categories of writing errors that generally are discussed are surface errors and global errors. According to Burt and Kiparsky (1974):

Global mistakes are those that violate rules involving the overall structure of a sentence, the relations among constituent clauses, or, in a simple sentence, the relations among major constituents. Local [surface] mistakes cause trouble in a particular constituent, or in a clause of a complex sentence. These are relative notions; something that is global in one sentence may become local when that sentence is embedded in a bigger sentence. (Burt & Kiparsky, 1974, p. 73)

In particular, evidence of local mistakes can be found in agreement, articles, and noun phrase formation in composing (Tomiyana 1980).
According to Zamel (1985), instructors generally focus on surface errors, which involve usage, style, and mechanics, rather than on global errors, which are concerned with wider issues of expression, in responding to student writing. Instructors seldom require writers to revise subsequent drafts of a paper beyond the surface level. Most ESL textbooks, Zamel says, stress the accuracy of surface-level features of writing rather than global issues. Yet, according to Burt and Kiparsky (1974, p. 79), only second language students commit global errors in their writing, indicating that ESL students do have a need for instruction in writing issues beyond the sentence. This is only one thought, although others do not hold this view.

Such a shift from global mistakes, which can affect the reader's understanding of a student's text, to surface errors offers "a very limited and limiting notion of writing . . . " (Zamel 1985) because students then tend to put their efforts into correctness. Thus, they may not learn the purpose of writing--the discovery of what they want to say. Focusing the stress of the teaching of writing on local mistakes (which in any event often do not obscure the meaning of the writer) rather than on global errors (which are more serious in nature because they can obscure the meaning of the text) does no great service to developing writers.
Additionally, there is another factor concerning which kind of error a writing teacher chooses to focus. Often teachers will address their comments to both the minor and major problems they see. Not only do beginning writers not know which mistakes or kinds of errors are most important, but they also find such comments confusing. For example, the students may find comments about local errors alongside comments regarding larger issues that either contradict the other comments or cause them to become unnecessary (Zamel 1985). One example of the latter type of comment is a teacher suggestion to delete a section of discourse which includes suggestions or comments for correction or change. Such feedback also does not promote the learning of writing as a recursive process of discovery.

Global Errors

As well as contradictory comments on student papers, instructors often write vague, overly-generalized comments that students are unable to understand or know what to do about. Sometimes teachers even write unclear notes regarding global issues (Zamel 1985).

Global matters, according to Burt and Kiparsky (1974), are more serious than are surface matters in writing (p. 73). Moreover, global errors are the easiest kind for students to appreciate and correct (p. 79). In a study of
nonacademic professional reader responses to student writing, clarity and economy were found to be valued above surface features (Hairston 1981).

Tomiyana (1980), in her investigation into error, discovered that
global mistakes typically confuse the relationship among clauses: use of connectors, distinctions between coordinate and relative clause constructions, parallel structure in reduced coordinate clauses, and tense continuity across clauses. (Tomiyana, 1980, p. 72)

Examples of the most typical global errors are terms and pairs used incorrectly, such as connectors ('and/but'), subordinating conjunctions ('because,' 'although,' 'if-then'), and the position of main and subordinate clauses (Burt & Kiparsky 1974, p. 79).

A Hierarchy of Error Gravity

In evaluating categories of errors, many researchers and teachers agree that there exists a stratification of writing mistakes. A handbook by Burt and Kiparsky (1972, p. 5) recommends that, rather than correcting every error, writing instructors use a hierarchy to which only the most grievous errors, those which most interfere with reader comprehension and communication, are attended. Such a taxonomy does indeed appear to be what most teachers formulate in evaluating writing. By way of illustration, an inquiry done by Vann et al (1984) discloses that, in faculty
response to NNS writing, most respondents do indeed appear to construct a hierarchy of errors.

Global mistakes do, indeed, seem to be the sphere of errors that are generally deemed the highest on the hierarchies of most teachers (Burt & Kiparsky 1974). Less agreement occurs, however, relative to the components and ranking of such a hierarchy. To exemplify this, the research done by Vann et al (1984) indicates that respondents rate most seriously the error typology characteristic of NNS's.

Vann et al list word order, 'it-'deletion, tense, relative clause errors, and word choice as the most serious errors. Notwithstanding their discovery, Zamel (1985) denotes that their study was comprised of only isolated sentences, not larger blocks of discourse wherein errors are set in context, as is genuine writing.

By contrast to the above error list, Hairston (1981) contends that the list should begin with non-standard verb use, double negatives, and beginning a sentence with an object pronoun (for instance, "Him and Sally used to be my friends"). Chastain (1981), on the other hand, finds that the most egregious errors occur in noun phrases and verb phrases. Nevertheless, he submits that even with the presence of these errors of utmost gravity, very few readers are unable to comprehend the message of the writer. Further, he reports:
These data support the hypothesis that some linguistic errors are more serious than others from a communicative point of view. At the same time, those errors that are understood and considered acceptable by native speakers could temporarily be ignored, from the strictly communicative point of view. (Chastain, 1981, pp. 293-94)

In contrast to the above lists of errors for ESL writing, NS teachers for English as a foreign language (EFL) writing find verb tense and concord errors the most serious, while NNS teachers feel the most prominent are case and lexis errors (James 1977). James, though, judges the principal error types to be transformations, verb tense, concord, case, negation, articles, and order. Moreover, he holds that the least serious kinds of error are lexical.

Further examinations regarding error gravity yield additional findings. Zamel (1985), by way of example, posits that the most serious of the mechanical errors are sentence fragments, sentence run-ons, inappropriate capitalization, 'would of,' and lack of agreement. To Greek NNS’s, however, the most difficult error to understand in English papers they reviewed is a misspelling (Hughes & Lascaratou 1982). For ESL teachers, according to Zamel (1985), language-specific errors are of much more concern than they are for other teachers. Moreover, in a consideration of form vs. word-level errors, NS readers react more negatively to errors of form than do NNS’s (Chastain 1981).
Tomiyana (1980) explored the effects of syntactic errors upon written communication to discover the relationship between grammatical errors and communication breakdown. She discovered that incorrect article usage is easier to repair than are omitted and wrong-choice-type connectors. In other words, Tomiyana believes errors with articles are less serious than are some connectors.

Chastain (1980) conducted a study in which NS's rated native English-speaking students in intermediate Spanish classes. Word usage problems were rated either "comprehensible and acceptable" or "comprehensible but unacceptable" as a personal decision by some readers; thus, Chastain (1980) says, native speakers can often comprehend utterances that are linguistically quite corrupted phonetically, semantically, and grammatically . . . comprehension is most severely limited by word usage, the use of a wrong word or the addition or omission of words . . . thus, the forms of words seemed to be a much less important factor in the communicative process than the correct use of the words themselves. (Chastain, 1980, p. 210)

Additionally, the gravity of any error must be determined by its situational context, and word-form errors reduce grades (Carlson & Bridgeman 1986). Shaughnessy (1977) also concurs that "the static around some errors is greater than that around others" (p. 122).
Coherence

In addition to surface errors and other global problems ESL students have with academic writing, frequently the writing of many of these students contains problems with coherence and/or cohesion. The coherence of a piece of writing means its understandability and clarity of meaning. The coherence of a paper, then, means how the groupings of ideas, or paragraphs, hang together to make a logical whole. Frodesen (1991) states that coherence in writing

... is a multidimensional feature, achieved partly through text structure, but also through the reader’s perception of the text’s appropriateness in a specific rhetorical context. (Frodesen, 1991, p. xvi)

Thus, the writer must respond to the assigned task, meet the expectations of the reader, and follow the conventions of the rhetoric of Western expository writing. American conventions that are problematic in particular for ESL students, according to Frodesen (1991), are as follows: focusing on central points of the reading, maintaining consistent meanings for key concepts, clearly showing the relationships between parts of the discourse, and developing effective patterns of paragraphs (p. xvii).

Another study dealing with coherence reports an analysis of native English-speaking students in graduate
Spanish classes (Azevedo 1976). In the study, students committed semantic, lexical, and idiomatic grammatical errors because of disjunctions with the writers' language concepts. Students, then, would express a concept circuitously, using more words than were necessary, by using direct translation from English. In this manner, the writers were inventing new words, words which were understood in neither Spanish nor English.

Cohesion

By contrast to the logical sense of a composition is its basic understanding, or cohesion. To put it another way, there must not be confusion or uncertainty in the mind of the reader concerning pronoun reference, verb tense continuity, connectors that link clauses together, relationships between ideas presented, and so forth.

In her study, Frodesen (1991) found that ESL students made more lexical and grammatical errors, which contributed to a lack of cohesion beyond the sentence level, than did the NS students (p. 328). Nevertheless, "some nonpass essays responded admirably in meeting many of the reader expectations[;] [s]ome nonpass essays had clearly structured and context-appropriate thematic development" (pp. 334-35). It should be pointed out that Frodesen's study, useful that it is, made use of "ESL" designations on the papers of NNS
students, which easily could have slanted the ratings of their papers.

Yet another report (Halladay & Hasan 1976) promotes an analysis of cohesion for the teacher/reader of student essays that involves the naming of different categories. These categories are comprised of semantic relationships that the writer creates by way of his or her choice of grammatical structures and vocabulary (p. 303).

Stotsky (1983), on the opposite hand, argues that the scheme of Halladay and Hasan for analyzing lexical cohesion is inapplicable because it is derived primarily from an examination of samples of conversational and literary discourse. Rather, she says, they should have inspected samples of the style of writing most required of students in the academic setting--expository essay writing.

Additionally, Stotsky (1983) points out, "words contributing to cohesive ties in exposition tend to be literate words, i.e., words that are more apt to be seen than heard, written than spoken." This fact might account for the problems students from other language backgrounds often have with cohesion. These students often do not have such a literate vocabulary in English.

Another study, which examined the English writing of native Arabic and Farsi speakers (Evola et al 1980), found conjunctions, pronouns, and articles were the most difficult
cohesive devices for the students involved (p. 181). Despite such difficulties for ESL students, however, the study found that such a lack of cohesive devices does not correlate with overall writing ability. Thus, the basic message of an ESL student might be understandable to the reader even with some (or perhaps even many) surface and global errors that affect cohesion and, to some degree, coherence.

ASSESSMENT

Because of or despite the understandability and content of student compositions, teachers evaluate what they read. Often this assessment involves response to error. When professional writers (such as poets, novelists, or storytellers), university students, and people in the professions responded in a study to what they liked best and least about their own writing, though, none of the respondents mentioned syntax, grammar, form, or even style (Miller 1982).

Despite this, many researchers conclude that NS English composition teachers are strongly influenced by ESL errors when evaluating NNS essays (Fein 1980, Homburg 1984, McGirt 1984, Perkins 1990, Sweedler-Brown 1993a). Since instructors do and must evaluate the writing done by their students, it is necessary to examine the various methods of assessment being used by them. Some of the methods include evaluations of the above writing factors, while others do not.
Odell and Cooper (1980) report on four evaluative approaches: the General Impression scoring procedure, the Analytic Scale, the "relative readability" method, and the Primary Trait Scoring Procedure. While these researchers find strengths and weaknesses in each rubric, they state that the Primary Trait Scoring System is the only procedure based on current discourse theory.

**Holistic Scoring**

One specific method of assessment commonly used at the present in colleges and universities is the holistic scoring method. Holistic reading sessions involve rapid reading for an individual impression of the quality of the writing, by comparison with all other writing the reader sees on that occasion. Holistic reading is based on the view that there are inherent qualities of written text which are greater than the sum of the text’s countable elements, and that this quality can be recognized only by carefully selected and trained readers, not any objectifiable means. And yet study after study . . . has found that these conditions are unreliable, and that considerable effort must be expended to establish and maintain reliable judgment. (Hamp-Lyons, 1990, p. 79)

However, the reliability of holistic scoring has been questioned by some researchers (Hake 1986, p. 161). In an examination of why trained raters find consistent holistic scoring difficult (Barritt et al 1986), it has been found that raters apparently do not stick to judging the texts alone. It seems that raters feel compelled to construct in their minds an author of the piece they are reading. Barritt
et al (1986) found they "were trying to build consistency into students' texts [as they read them holistically] by investing spaces of indeterminacy in them with our own expectations about what should fill the gaps." When the expectations built up concerning the author are not met, then, such an atypical paper can easily cause inconsistencies in scoring responses among the raters. Such inconsistencies, if great, amount to unreliability. Since reliability is only a precondition for validity but not its guarantee, when reliability is lost, so is validity.

The Holistic Scoring assessment system does have a subjective component to it. Therefore, trained raters must agree on how to deal with great variations between mechanical and organizational abilities and must construct a hierarchy of error gravity (Carlson & Bridgeman 1986, p. 144; Homburg 1984). Holistic ratings often are swayed by content, organization, sentence structure, and mechanics, items which do not necessarily obscure the message to the reader (Carlson & Bridgeman 1986, p. 143). Therefore, when the writing competence of ESL students is examined, allowance should be made for their differing organizational structures and methods of expressing concepts (p. 126).

Even when readers do agree on a writing construct and a hierarchy of errors, however, scores can vary considerably among raters. Rhetorical differences such as patterns of
organization can serve as a "hidden agenda" that teachers untrained for ESL teaching hold; rather than a purported facility with English writing, they expect fluency (Land & Whitley 1989, pp. 284-85).

Hamp-Lyons (1990, p. 80), by contrast, finds the reason for variations in holistic ratings is that raters do not share the same construct of writing quality or, in other words, researchers "cannot consistently agree with each other when assessing the same writing samples or even sometimes with our own judgments about the same samples made on different occasions" (p. 80).

Still another finding focuses on writing topic assignments. Reid (1990), in her examination of how ESL students perform in response to various topic types, did not observe differing writing skills between two kinds of topics (Comparison/Contrast and Take a Position, and Describe and Interpret a Chart or Graph). In opposition to this is the study by Carlson and Bridgeman (1986, p. 148) which indicates that different writing topics elicit a variation of syntactic ability.

Ruetten (1991) states that pairs of holistic scores greatly at variance are found in ESL rather than NS writing. When the content is clear and developed in a NS essay, the grammar and mechanics are overlooked; ratings for NNS papers, however, include grammatical as well as rhetorical
features. Additionally, another study of rating pairs
(Mullen 1980, p. 167) revealed that some pairs could not
produce reliable judgments.

Moreover, such interrater unreliability can exist
across some language groups, according to Carlson and
Bridgeman (1986, p. 143). They cite a study conducted by
Breland and Jones (1982) in which Hispanic ESL students'
syntactic and lexicographic scores were much more important
than their scores for discourse characteristics.

Some researchers, while granting that holistic scoring
is highly subjective, still find it a reliable method
(Homburg 1984; Carlson & Bridgeman 1986, p. 149). For
instance, Homburg (1984) found that raters utilize several
features, one at a time and always in the same order and
manner, in scoring papers (a "funnel method"). Since the
categorization is comprised of a combination of features
such as T-unit length or number of errors, he finds
consistency and therefore objective evaluations.

One analytic investigation centered on which aspects of
the rating process raters stress (Freedman & Calfee 1983, p.
85). Of significance are the essay itself, the rater, and
the context, or setting, of the session. The investigators
show that the individual essay carries more weight among
expert evaluators (p. 77).
However, the above study does not address the issue of the validity of holistic scoring for ESL papers. Another study (Kroll 1983) does; it discovered that holistic scoring was able to provide both reliable and valid scores (p. 27). Similar scores for each student were recorded for papers written at home and those produced in class (p. 62). However, Kroll notes that ESL students still produced erroneous word level choices for at-home compositions (p. 141).

A COMPARISON OF NATIVE- AND NON-NATIVE WRITING

Some studies propose that the writing problems of NS and NNS students are the same or similar enough for both groups to be taught in the same classroom. Amberg (1984), for example, found that developmental-level NS's and advanced-level ESL students score similarly on writing tests. In addition, the two groups of writers committed the same kinds of errors: subject-verb agreement and verb usage (choice, tense, and verb form errors).

On the opposite side, Cohen (1975, p. 196) found there are notable characteristic differences between the errors of NS's and NNS's. In addition, Benson et al (1992) found the two groups should be separated because non-ESL-trained teachers are not equipped to deal with the cultural and writing differences of ESL writers.
One more reason for separating classes comprised of both student populations is that NS’s and NNS’s vary in their strengths and weaknesses in two distinct categories of grammatical facts, according to Nattinger (1978): hard and soft facts. Hard facts, for instance, are

- externally perceivable events which are operationally definable; they usually involve measuring and counting, and are independently verifiable by the replication of some sort of procedure; they are the facts of what humans do. The other kind of facts are "soft facts." These are more internal responses to the external data; they are only internally verifiable and are perceptions of what humans think they do; they involve, in other words, the mental concepts with which we categorize the world. (Nattinger, 1978, pp. 77-78)

While NS’s share the same soft, or culturally-related, facts with NS instructors, NNS’s often share neither the hard facts nor very many of the soft facts with their teachers.

The most common difference in writing that ESL student papers present, though, is what appears to be redundancy (Land & Whitley 1989). Nevertheless, readers can reread an essay until they recognize the structure utilized by the writer. If a reader utilizes a topical structure analysis, rereading and noting all grammatical subjects of independent clauses, the reader often sees that these subjects operate as higher-order cohesive devices (p. 291).

By contrast, Grabe and Biber (1987) reached a different conclusion when they performed a textual analysis on the relationship between the co-occurrence of various patterns,
such as structural features, analyzing the genres of narrative, exposition, argumentation, and description. They applied a multi-feature/multi-dimensional approach.

Grabe and Biber conclude that there is almost no difference between the papers composed by NS and NNS student writers. Furthermore, they found only small differences between essays rated "good" and "poor." However, these researchers conclude that ESL students utilize more pronouns than NS's, indicating a less-formal academic style.

A similar study mentioned earlier (Reid 1990), however, obtained results at odds with those of Grabe and Biber. Reid used a discourse analysis to compare the writing of NS's of four languages to identify quantitative differences in the syntax and lexicon of the languages. The investigation consisted of four writing tasks, two topic types and two topic tasks for each type (Comparison/Contrast and Take a Position, and Describe and Interpret a Chart or Graph). Reid discovered different results than Grabe and Biber concerning excessive use of pronouns. Students respond by topic type in their use of pronouns rather than using them excessively for all topics and types (p. 202).

COMPARING SCORES BEFORE AND AFTER CORRECTION

Thus far, we have seen that ESL students generally commit more and usually distinctly different errors in their
writing than do NS students. The preponderance of local errors in particular appears to affect the scores received by NNS compositions. Moreover, ESL writers often compose using their first-language rhetorical patterns, which vary from what is expected in American institutions of higher learning. Thus, the grades of NNS’s often are lower than their NS classmates.

What this research will attempt to discover is whether factors other than surface errors, such as rhetorical patterns that vary from typical American coherence, cohesion, and so forth, affect the scores of NNS compositions. Several researchers have investigated scores for comparable groups of students from the two populations, and a few more currently have reported on comparisons of scores before and after correction for NS’s and NNS’s.

Several researchers, for example, state that higher grades are given to NS papers than to error-free, NNS papers (Land & Whitley 1989, p. 286; Fein 1980; McGirt 1984; Whitley 1984; Sweedler-Brown 1993a). However, in an inquiry into whether the compositions for the two groups in comparable courses are rated similarly, Fein (1980) discovered systematic differences in the writing.

Conversely, it is noteworthy that some measures in his inquiry do not indicate gross disparities between the groups with ESL essays scoring lower. For instance, NNS papers were
comprised of about as many words as those of NS’s, had fewer
discourse-type errors, wandered off the topic less
frequently, and contained slightly higher scores in mean T-
unit length (indicating that the ESL writing samples may
have been more sophisticated).

Then, in an attempt to determine the effect of
sentence-level errors on the ratings of ESL papers, one
researcher (Sweedler-Brown 1993a) compared scores for before
and after correction of ESL students within a control group
of developmental-level NS’s. She found significant score
increases for the mean of the lower- and higher-rated ESL
compositions. Curiously, the mid-range scores did not
increase similarly. Moreover, while the error rate before
correction was high (8.99/00 words) for the strongest ESL
essays, they did receive passing scores. The assumption is
that raters were influenced more by variables such as
rhetorical variation or weakness rather than by surface
features in the writing.

Next, McGirt (1984), Whitley (1984), and Sweedler-Brown
(1993b) expanded the topic of score comparisons by
conducting studies that contrasted ESL scores before and
after correction with those of NS papers. While McGirt
(1984) compared samples taken from comparable groups that
fulfill the undergraduate composition requirement at the
University of California at Los Angeles, Sweedler-Brown
(1993b) compared ratings derived from essays produced by intermediate ESL students and developmental-level NS students at San Diego State University.

McGirt (1984) determined that the essays composed by 40% of the ESL students in his study were impaired by the frequency of morpho-syntactic writing errors. In addition, such surface-level errors are not the sole reason for the unsatisfactory ratings of the ESL papers. On the other hand, by examining analytic scores of various writing factors as well as those of holistic scoring, Sweedler-Brown (1993b) discovered that sentence-level errors alone accounted for the significant variations in ratings for before and after correction in her study. A notable finding was that of the pass/fail difference between the two groups: 16 received failing scores and 2 received passing scores originally; after corrections, 1 received a fail and 17 received a pass. Thus, in her study the presence of surface errors is extremely significant for the successful completion of the course by ESL students.

All of the studies above indicated in mean comparisons that NNS scores before correction were failing. Moreover, after correction, the means for all were a strong pass. The t-tests all indicated significant differences between the means of the two populations. Moreover, all groups began with relatively high percentages of failing scores and
concluded with notably lower percentages. Next, the mean error count for each study indicated dramatic differences before and after correction. Finally, all studies indicated dramatic variations in comparison with the NS scores, and in each instance the NS scores did not increase as dramatically after corrections were made.

The findings from the literature discussed contain much disagreement concerning both the written efforts of ESL students and instructor response to those efforts. In particular, there exists a lack of sufficient studies of comparisons between NS and NNS texts in relation to surface errors. The aim of this study is to add to the efforts of other researchers regarding this issue.
CHAPTER THREE

THE STUDY: SUBJECTS AND METHODOLOGY

CONTEXT

For the purpose of clarifying the context of this study, I will summarize the curriculum and placement procedures for the composition courses offered at California State University, San Bernardino. Following that, I will then discuss the background of the subjects of the study. I will then describe the procedures of the study. In this way, the various parts of the study can be placed in their proper perspective.

Incoming undergraduate students at California State University, San Bernardino must take the English Placement Test (EPT). Placement in freshman composition or one of several preceding basic writing courses depends upon the score received on the EPT. The courses prior to freshman composition, English 101, begin with English 85A, then proceed to English 85B, and conclude with English 95. These courses carry credit, but they are not degree-applicable. Students can move up through the ranks of these courses by successful completion of them, or they can be recommended by their teacher(s) to be placed directly into 101 from 85A or 85B.
During the winter 1992 term, the university offered two comparable freshman English courses, English 101 and ESL English 101. Both courses met the undergraduate writing requirement, and both received the full credit of a university course, four units. Only one section of the ESL English 101 course was offered, as a pilot course, and only during that one term.

The procedures for placement for that course are the following: early in the fall term of 1991, a writing prompt was administered for all winter 1992 term incoming freshman composition classes. The students were also required to complete a biodata form, which requested information concerning their native language and background. This form is reproduced in Appendix A.

Essays were then pulled and listed as appropriate for ESL English 101 according to two criteria: the student's self-report, which noted that he or she was a non-native-speaker of English and/or the writing's exhibiting any ESL features. This determination was made by the TESL specialist, Professor Wendy Smith, and by two other experienced ESL instructors.

Then, individuals from the list of NNS students were invited to participate in a pilot ESL English 101 class. The subjects of this study came from one section of an English 101 class and the above-described ESL English 101 class.
SUBJECTS

The study is comprised of thirty-four students, twenty from English 101 and fourteen from ESL English 101. The instructor for the English 101 class reported that there were no students in the class who had a background of L1 (i.e., first language) other than English and who had resided in the United States for less than seven years. Three papers had been excluded from the study for reasons mentioned below (see "Procedures"). There were twenty-three females and eleven males remaining in the pool of eligible subjects who participated in the investigation. Thirteen of the females were N’s; ten were not. Of the males, seven were NS’s and four were NNS’s.

A wide diversity of language backgrounds is represented among the subjects participating in the ESL sample. Below is a list of the native languages of the ESL students:

<table>
<thead>
<tr>
<th>Language</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amharic</td>
<td>1</td>
</tr>
<tr>
<td>Cambodian</td>
<td>1</td>
</tr>
<tr>
<td>Farsi</td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td>1</td>
</tr>
<tr>
<td>Laotian</td>
<td>1</td>
</tr>
<tr>
<td>Malay</td>
<td>1</td>
</tr>
<tr>
<td>Mandarin</td>
<td>4</td>
</tr>
<tr>
<td>Spanish</td>
<td>4</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>3</td>
</tr>
</tbody>
</table>

PROCEDURES OF THE STUDY

The research for this study involves comparisons of written work produced under comparable conditions by the
students in both a NS and a NNS freshman composition class. The study was designed to address the following questions:

1. Is there more variation in holistic scores awarded to ESL-written essays than to native-speaker essays, in other words, less inter-rater reliability?

2. With surface errors corrected, do the ratings increase comparably for both groups?

3. Do raters respond to surface errors more or to some other non-native quality in the writing, as judged by their response to corrected, error-free compositions?

THE ASSIGNMENT

In-class compositions were elicited from the students during a midterm based on a common reading/writing assignment. The topic for the current study was agreed upon by the participating instructors before the testing occurred (Appendix B). The text which was used for the prompt was an article entitled "Friends as Family" by Karen Lindsey (Colombo et al, eds., 1992, pp. 463-476). The assignment occurred during the fifth week, or approximately halfway through the winter quarter, on February 13, 1992.

Teachers notified students approximately two weeks in advance that they would be writing an in-class composition on the article. Both instructors opted to discuss the article and possible prompts regarding it during a class meeting before the day of the common writing. The article
was distributed to the students a week before the session allotted to the in-class composition.

Students were not allowed any outline, notes, or other materials other than a copy of the article and a dictionary during the course of composing. Discussion concerning neither the article nor the prompt was permitted during the class session in which the writing occurred. Sixty to seventy minutes were allotted by all instructors for the in-class writing. None of the students whose papers were involved in this research were aware of the study being conducted.

SELECTION OF THE DATA BASE

Next, from the two classes whose instructors had agreed to participate, I examined the compositions for their appropriateness for the study. Papers from three of the NNS students were excluded from the study at this point for various reasons. In one instance a writer referred, in the body of her composition, to her own name. Such comments could indicate to raters that the individual might be a NNS. Moreover, two more students were dropped from the study for specifically stating in their essays that they were foreign students studying in the United States.

I then typed the handwritten essays for the thirty-four participants as they had been written, leaving intact any
and all errors. Names were not included on the compositions being scored; rather, I assigned numbers to each paper. Moreover, care was taken to avoid alerting the raters of the NNS status of any of the student writers. Such action was necessary in order not to distort the study; the possibility existed of rater variation in response to perceived ESL texts. Such response could have elicited different reactions. See Appendix D for samples of essays before and after correction for both groups of writers.

SCORING

The papers were then presented to the freshman 101 committee for scoring. No teachers evaluated the writing of their own students. The scoring of each composition was performed by two English department normed raters; i.e., they had been trained by the department to score compositions holistically during a rapid reading session. Additionally, the raters were experienced in such holistic scoring sessions.

Pairs of normed raters in the English department at CSUSB generally are able to assign scores, on a six-point scale, within two points of each other. In the event of a difference greater than two points, however, papers are given to a third rater. The score assigned is then the average of the two closest scores.
The six-point scoring guide used for the holistic rating sessions pertaining to this study is customary for the English department, in that all teachers using the guide for holistic scoring have been trained in and are experienced with the guide. See Appendix C for a copy of the rating scale used by the English department.

IDENTIFYING ERRORS

After the scoring was completed, I compared each set of scores. The next step was to examine all papers for morphosyntactic errors (see Appendix E for the error list used). For this study, I used the same categories of errors as those used in the comparable study done by McGirt (1984). The error list originates from a pilot study conducted by ESL and TESL (teachers of English as a second language) consultants. The purpose of the study was to form a basis on which McGirt (1984) could develop a policy for identifying and correcting errors. Mechanical, word-level, and sentence-level errors are included in the list, which appear below. I found no errors in my data base other than those categorized in the list.

CORRECTION OF ERRORS

These errors were corrected in the next typing of the compositions. The error types were of items that were
omitted, superfluous, or of the incorrect form. The error correction guidelines which were developed by McGirt (1984) exclude errors that are typical of NS writing, such as faulty parallelism.

The second item to note is that examples of errors of morphological and syntactic origin which were corrected are agreement, articles, function-word choice, local word order, word redundancy, missing word, and lexical-syntactic problems. Examples of these errors as exhibited in NNS essays are as follows.

1. Agreement:
   The article . . . is the perfect examples
   Everyone expect to have a perfect, happy family

2. Articles:
   [The] Biological family is important
   There are strong bonds among [a] family's members

3. Function-Word Choice Errors:
   She talked about the adopting family
   the nuclear family, which carries the blood bondage

4. Local Word Order Errors:
   some changes about families' model from '50's
   a perfect, well-rounded, run smoothly family.

5. Redundancy:
   You have your own mind and self, decide for
yourself. The traditional family is up to your
decision . . .
We can choose most of our family; we can choose
all of our family. Her opinion about family . . .

6. Missing Word:
First [__], "honorary relative" family that uncle
or aunt as you grow up.
but it is not the most popular [kind of] family.

7. Lexical-Syntactic Problems:
The first family that I would want to
mentioned . . .
They hate each other and love is not existing.

Third, mechanical errors to which I attended were
spelling, capitalization, and punctuation. Some examples of
these errors follow.

1. Spelling:
   society [society].
asolute [absolute].

2. Punctuation:
   She did not believe in the _myth of nuclear
   family.
   The nuclear family is more than a myth[_] a lie.

3. Capitalization:
   she describes how she feels about certain
   families.
the conditions for calling a group of people a family are simple.

Specifically, the major grammatical categories which were used for corrections are as follows:

1. Agreement
2. Articles
3. Conjunctions
4. Derivational Morphemes
5. Fragment/Run-On/Comma Splice
6. Mechanics
7. Prepositions
8. Pronouns
9. Singular/Plural
10. Verbals
11. Verbs
12. Word Choice
13. Word Order

In addition to the above list, some of the major grammatical categories, as noted in Appendix E, contain several sub-categories. For instance, pronouns may be erroneous in case or reflexiveness, and verbs can contain errors in form, passive voice, auxiliary, modal, or particle (a loose affix). Moreover, derivational morphemes are broken down into nominal or adjectival forms. In this list, however, adverbial forms of derivational morphemes have been omitted.

Punctuation, furthermore, contains eight sub-categories. Spelling errors are categorized in four manners, as are conjunctions. There are both '-en' and '-ing' listings for participial phrases, and three categories of
singular versus plural nouns, which are count/noncount, noun adjuncts, and object of preposition.

SECOND READINGS

I next typed the essays in the corpus, correcting all of these errors. No other changes were made in the compositions. As with the papers before correction, I left off student names and instead assigned numbers. Thus, the identities of the authors of the papers before correction that were included in the study and the papers after correction were unknown to anyone other than this researcher.

After that, the papers were given to other normed raters from the English department who were experienced with holistic scoring procedures and the campus scoring guide. No rater received both compositions before correction and after correction. The readers were given no explanation other than that I wanted them to use the above-mentioned six-point scoring system to evaluate the compositions holistically for an unspecified study I was doing. These raters were similarly uninformed of the inclusion of ESL papers in the corpus or of the purpose of my research.

Following the second rating of the essays, the mean and standard deviation were determined for the four groups: NS
uncorrected, NS corrected, NNS uncorrected, and NNS corrected.

Next, correlation coefficients were performed to determine the reliability of the raters. Last, after collecting the second set of rated papers, statistical analyses (t-tests and MANOVA) were performed on both their results and on those from the essays before correction. The t-tests served to compare the means of the two populations, and the MANOVA served to determine the significance in score variations. The two sets of results were then compared. The findings are described in the following chapter.
CHAPTER FOUR

RESULTS

The holistic scores assigned by the raters to the essays from the two groups of writers, NS and NNS, reveal both expected and unexpected results. The scores before correction are lower than the scores after correction for both groups. In both instances, NS students as a whole receive higher ratings than do NNS students. These results concur with those found by other researchers (McGirt 1984, Whitley 1984, Sweedler-Brown 1993b, Fein 1980).

However, a closer examination indicates scant difference in comparing the amounts of increase in ratings from the versions before correction to the versions after correction. These statistics are discussed later. This result differs from the similar studies conducted by the above-noted researchers. However, as I expected, interrater reliability decreases significantly for NNS scores after correction.

INTERRATER RELIABILITY

An examination of the responses of the pairs of ratings for the two groups of papers yields unexpected results. One result is that the scoring sessions produced what appeared
to be agreement in response among the raters. However, correlation coefficients, which reflect statistical reliability, indicate less agreement than initially appeared.

For the scores before correction, the figures indicate a high positive correlation between pairs of raters for both groups of ratings. The figures are +.721 for NS’s and +.73 for NNS’s. In other words, much reliability exists before correction between the pairs of raters for both groups.

On the other hand, correlations decrease for both groups on scores after correction. The NS figure now is .624, while the NNS slips considerably to .574. Thus, not as much agreement (and therefore not as much reliability) exists after surface error correction, particularly for the NNS population. The response to the first research question, then, regarding whether variation in holistic scores between the two groups of writers exists, is positive: I conclude there does exist notably less agreement between raters on rating NNS essays without surface errors.

One peculiar result is that of differences in scores for pairs of raters before correction and after correction. The ratings for 12% of all papers actually decrease from before correction to after correction. The reason for such a drop in ratings might be explained by variations between rater response pairs.
TESTING PROCEDURES BEFORE CORRECTION

The mean and standard deviation for the uncorrected essays were calculated. These statistics are indicated in Table 1. Before correction, the mean is 6.60 for NS’s and 6.07 for NNS’s. For NS’s the standard deviation is 1.39; for NNS’s it is 1.10. I discuss these statistics later.

Descriptive Statistics
for Scores of
Uncorrected Compositions

<table>
<thead>
<tr>
<th>GROUP</th>
<th>$\bar{X}$</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>6.60</td>
<td>1.39</td>
</tr>
<tr>
<td>ESL</td>
<td>6.07</td>
<td>1.10</td>
</tr>
</tbody>
</table>

TABLE 1

Table 2, a list of the frequency distribution of scores, indicates that 43% of the NNS papers score at or below the lowest score obtained by NS papers, 5.0. Likewise, 20% of the NS papers score at or above the highest score obtained by NNS papers, 8.0.
Frequency Distribution of Scores
for Uncorrected Essays

<table>
<thead>
<tr>
<th>Score (f)</th>
<th>Score (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>9.0</td>
<td>9.0</td>
</tr>
<tr>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

**TABLE 2**

The results from the frequency distribution of scores are presented visually in Figure 1, Uncorrected Scores. This graph shows the percentage of each group of students that receives the various scores. The figure charts an unusually-shaped line for each group: it progresses in a mostly upward direction rather than forming the usual "teepee" shape similar to the typical bell-shaped curve. There is a near overlap between the two groups at about 20% between a one-point spread of scores. Both the NS's and NNS's appear to fall within several distinct groups.

In comparing the table with the information in the frequency distribution chart with mean and standard deviation (Descriptive Statistics, Table 1), we see that 36% of the NNS students receive ratings above the mean of the NS students. Further, 50% of the NS's receive scores at or below the mean of the NNS group.
Generally, scores for the two groups of students appear to be distinctly separate. In order to determine whether the difference between them is significant, a t-test was conducted.
RESULTS OF THE T-TEST

The t-test performed on the calculations before correction yielded mixed results. Specifically, the findings are as follows: the results indicate that score variations between the two groups, NS and NNS, are such that they do not occur by chance. In other words, there is a significant difference between the NNS and the NS scores before corrections were made. In contrast to this finding, there is no significant difference between the means for the corrected NS and NNS ratings.

TESTING PROCEDURES AFTER CORRECTION

I then performed similar operations on the results of the ratings after correction. Both sets of scores were recorded; each pair of ratings was totalled. No two raters for either group deviated more than two points.

The next step was to average the pairs of ratings for all four sets of scores. Several interesting findings emerge. One finding is that large percentages of papers would have failed before corrections were made, according to the six-point grading scale, since only papers scoring above three points pass. 55% of the NS's and 64% of the NNS students received non-passing grades before corrections were made (see score descriptions in English Department Rating Scale, Appendix C).
Scores after correction indicate interesting results also. Most scores for both groups increased; only 25% of the NS's and 36% of the NNS students received nonpass ratings. For the NNS papers, 36% did not increase their scores, 7% lost points, and 29% raised their scores from failing to passing. No papers slipped from a pass to a fail rating.

Descriptive Statistics
for Scores of
Corrected Compositions

<table>
<thead>
<tr>
<th>GROUP</th>
<th>( \bar{X} )</th>
<th>( s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>7.35</td>
<td>1.49</td>
</tr>
<tr>
<td>ESL</td>
<td>6.86</td>
<td>1.12</td>
</tr>
</tbody>
</table>

**TABLE 3**

In Table 3, the descriptive statistics for corrected compositions, the mean and standard deviation are seen as 7.35 and 1.49 respectively for NS's, and 6.86 and 1.12 for NNS's respectively. Compared to the figures before correction, these all show an increase. The increase for NNS standard deviation indicates the point spread is larger, signifying that less agreement in scoring response exists.
Frequency Distribution of Scores for Corrected Essays

<table>
<thead>
<tr>
<th>Score (f)</th>
<th>Score (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>9.0</td>
<td>9.0</td>
</tr>
<tr>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

TABLE 4

I again computed the frequency distribution of scores (noted in Table 4). A comparison with the similar table before correction yields an overall increase in scores for both groups, as might be expected. In contrast, though, this scoring session yields little difference in ratings between the two groups. This time, only 14% of the NNS essays score at or below the lowest score obtained by NS papers, 5.0. Moreover, 20%, the same amount as before, of the NS essays score at or above the highest score obtained by NNS papers, 9.0.
The visual presentation of these figures offers a dramatic difference in that, rather than the prior incline, the percentages now form a more familiar, up-and-down,
triangular design (see Figure 2). Similar to the figure representing the papers before correction, this figure depicts two separate groups, yet they now appear to overlap to a greater degree. Moreover, now only for the NS’s are there several one-point spreads of identical percentages.

A comparison of the figures in the table of descriptive statistics with the frequency distribution chart shows that approximately 64% of the NNS essays now receive ratings above the mean of the NS essays. Moreover, 45% of the NS’s now receive scores at or below the mean of the NNS essays. Thus we see that, after correction, the scores for the NNS compositions have increased dramatically in proportion to the NS compositions, when viewing the data this way.

Mean Score of the Four Groups

<table>
<thead>
<tr>
<th></th>
<th>With Errors</th>
<th>Without Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6.60 (NS)</td>
<td>7.35</td>
</tr>
<tr>
<td>9</td>
<td>6.07 (ESL)</td>
<td>6.86</td>
</tr>
</tbody>
</table>

FIGURE 3
Furthermore, a comparison of the mean scores makes the picture visually and conceptually clearer. Figure 3 indicates the means before and after correction for both the NS and NNS ratings. As can readily be seen, the scores for the two groups of students plot slightly upward, nearly parallel, on inclining lines. This depiction indicates the similarity in rating increase for both NS and NNS papers.

**RESULTS OF THE T-TEST**

The results of the t-test conducted on the scores after morpho-syntactic error correction indicate there is no significant difference between the means for the uncorrected and corrected NNS sample.

Because the results of the above statistical calculations were somewhat conflicting and indeterminate, a MANOVA (multivariate analysis of variance and covariance) was then performed on the computations.

**RESULTS OF THE MANOVA**

This study was designed with two independent variables: native status/non-native status and errors/no errors. A MANOVA was performed for repeated measures, that is, to determine whether the slight difference in scores between the two student populations could account for a variable or for the interaction between the variables.
### MANOVA Tables

#### Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>Sig of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Cells</td>
<td>90.94</td>
<td>32</td>
<td>2.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2974.83</td>
<td>1</td>
<td>2974.83</td>
<td>1046.79</td>
<td>.000</td>
</tr>
<tr>
<td>Group</td>
<td>4.30</td>
<td>1</td>
<td>4.30</td>
<td>1.51</td>
<td>.228</td>
</tr>
</tbody>
</table>

#### Tests Involving "SCORE" Within-Subject Effect

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>Sig of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Cells</td>
<td>27.05</td>
<td>32</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>9.71</td>
<td>1</td>
<td>9.71</td>
<td>11.49</td>
<td>.002</td>
</tr>
<tr>
<td>Group by Score</td>
<td>.01</td>
<td>1</td>
<td>.01</td>
<td>.01</td>
<td>.938</td>
</tr>
</tbody>
</table>

TABLE 5

The figures for the MANOVA appear in Table 5 above. One can conclude from the MANOVA that, for both groups, the statistics for corrected essays are significantly better than for the scores for uncorrected essays. There is no significant main effect due to group; in other words, the NS and NNS students have performed similarly in this study. The improvement from uncorrected to corrected score is the same for both NS and NNS ratings. The table indicates no statistically significant interaction between student status.
(NNS or NS) and error (with or without errors). From these results, we see there is no significant rating variation that is attributed to either +/- nativeness or +/- error. In the following chapter, I discuss these results.
DISCUSSION, SUMMARY, AND SUGGESTIONS

CHAPTER FIVE

DISCUSSION OF THE RESULTS

This study examines the effects of morpho-syntactic errors on the writings of NS and NNS essays. The study compares the holistic ratings assigned to each group of papers. The results of the research are as follows:

1. For holistic scoring of ESL texts, pairs of raters are generally as reliable as for native texts before correction; after correction, there is considerably less agreement.

2. Without the control of differences between the morphological and syntactic errors, the teacher response to NS and NNS writing is significantly different.

3. With the control of morphological and syntactic errors, instructor response to the writing proficiency of NS and NNS writers does not exhibit a statistically significant difference.

4. The scores after correction of morpho-syntactic and mechanical errors does not exhibit a significant difference between the ratings of native or non-native compositions.

This chapter discusses the results of the study and the writing proficiency of the NS and NNS subjects who participated in the study. Additionally, the chapter reviews the reliability of holistic scoring, specifically of NNS writing, for this research. Further, I include a summary and conclusions of the research, followed by some implications.
to which the findings point. Lastly, I offer several pedagogical suggestions regarding the evaluating of ESL student writing.

RESULTS: RESEARCH QUESTION 1

Is there more variation in holistic scores awarded to ESL essays than to NS essays, in other words, less interrater reliability?

The apparent agreement exhibited between the pairs of raters during holistic scoring was unexpected. That is, most researchers (Carlson & Bridgeman 1986, Homburg 1984, Mullen 1980, Hamp-Lyons 1990, Hake 1986, Ruetten 1991, Land & Whitley 1989, Freedman & Calfee 1983, Janopoulos 1989) have shown that, in particular, the holistic scoring of NNS compositions often produces unreliable pairs of ratings. I expected that ESL-type, beyond sentence level, rhetorical differences would have created less reliability than the NS ratings, requiring third readers in many instances.

The scores for pairs of readers, however, were in accordance more than anticipated. This finding might be the result of the small size of the sample. Another possibility for such concordance in pairs of ratings is that the raters at CSUSB, although not all trained in the field of ESL studies, have received excellent training in holistic scoring such that they can attain a high rate of agreement in their scoring. On the other hand, it is possible that
these raters all attend so much to surface error that that is all they see.

One questionable point concerning the holistic scoring in this study is the decrease in some scores for both groups after correction. Overall, though, from an examination of the ratings there appears to be more interrater agreement here than previous research would suggest. Despite the above findings, though, the results of the correlation coefficients are more revealing of the raters' responses. While the correlations before correction are quite high for both groups, the decreased correlations after correction indicate less agreement. This is particularly the case for teacher response to NNS writing efforts.

For this reason, we see that agreement exists for these raters, in the main, in regard to response to surface errors. However, there appears to be less agreement in response to NNS composing per se, that is, after surface errors have been eliminated. In this manner, this study indicates the raters are in agreement in response to surface errors and attend to them to a degree greater than they claim.

RESULTS: RESEARCH QUESTION 2

With surface errors corrected, do the ratings increase comparably for both groups?
The raters of the second group of compositions, those which had morpho-syntactic errors corrected, rated the essays of both NS’s and NNS’s higher than the first scoring before correction. Such a result is to be expected. These raters awarded scores that were different for the two groups of papers—the NNS essays received lower ratings overall. In spite of the difference between the performance of the two groups, though, the difference between NS and NNS after correction was not significant. This result is surprising; I expected a negative answer to this question, one that resulted from a dramatic increase in NNS scores, yet a mean significantly lower than for NS scores. I elaborate this point in my response to the following research question.

RESULTS: RESEARCH QUESTION 3

Do raters respond to surface errors more or to some other non-native quality in the writing, as judged by the response to corrected, error-free compositions?

From the multiplicity of studies done by other researchers, I expected this study to conform to their findings, that raters of holistic scoring would rate NNS papers, even after surface-error corrections had been made, significantly lower than NS papers. Nevertheless, such was not the case in every manner.

In this study, although NNS compositions are scored lower than the NS’s before correction, they are not after
correction. In fact, a smaller difference exists after correction between the NS and NNS scores.

For this reason, it appears that the raters in this study did not respond to any factor in the writing of NNS students differently than they did to the NS papers. In other words, at least some raters appear to respond solely to surface errors in some or all NNS essays. Some of the categories of surface errors exhibited in these essays are agreement, article, derivational morpheme, punctuation, word order, and spelling. Additionally, the scores for the NNS essays did not increase dramatically from uncorrected to corrected. Therefore, I conclude that the raters in this instance appear to have responded mainly to the initial surface errors.

The scoring guide (Appendix C) specifically states for raters to consider overall quality of the writing. Also, the descriptions for unsatisfactory scores describe weakness of focus and/or structure. Most of the literature for response to writing describes lower scores for NNS essays because of coherence, cohesion, and other rhetorical problems. Therefore, I further conclude that the raters for the majority of the papers for this study did not attend as much to these larger issues as they did to sentence-level features, to which they responded vigorously.
WRITING PROFICIENCY OF NATIVE SPEAKERS

In this study, initially over half of the NS compositions (55%) receive failing marks. After error correction, though, the scores increase appreciably, causing some failing essays to pass. Nearly half of the papers with failing scores now receive passing scores. Nonetheless, the scores of several essays do not increase. In comparison with the NNS papers, we see that only about half of the NS papers (55%) are rated at or below the NNS mean after correction, as opposed to 20% before correction.

Overall, the performance of the NS students in this study offers few surprising results. However, I did not expect this pool of writers to perform as weakly overall as they did.

WRITING PROFICIENCY OF NON-NATIVE SPEAKERS

This research indicates that NNS writers commit a substantial amount of morpho-syntactic errors of the category generally identified as ESL-type errors. In turn, such errors cause the scores of these students to be downgraded considerably. While over half of the NNS essays, for instance, initially receive failing scores, their scores after correction increase dramatically: only about a third now receive failing scores. Nevertheless, a considerable amount of the compositions still do not increase their
ratings after error correction. In these instances, I assume that issues other than surface errors affect the ratings. For example, such issues could be coherence, cohesion, or other rhetorical features. Many of the NNS writings receive scores at or below the lowest score obtained by NS papers before correction, while a much lower amount score at or below NS writers after correction.

Thus, surface errors do seem to affect NS raters in holistically scoring NNS compositions. However, after correction of minor morpho-syntactic errors that do not obscure the meaning of the writer’s message, this student population increases scores at a rate similar to NS’s.

The latter result of this study certainly is unexpected--all of the previous research strongly suggests that variations of rhetorical patterns for ESL writers influence ratings negatively. Thus for this study, some raters did not respond in all cases to matters other than surface errors, such as weakness or rhetorical variation in the writing of NNS students. Moreover, these raters did not increase the scores of the NNS papers at a rate higher than that of the corrected NS papers.
CONCLUSION

SUMMARY

Fein (1980), McGirt (1984), Sweedler-Brown (1993b), and Whitley (1984), in their research, found that NS students score higher on holistically rated compositions than NNS students, both before and after correction of morphosyntactic errors. The work conducted by Sweedler-Brown (1993a) also supports their research, finding that some ESL writing is awarded significantly higher scores after error correction as compared to before correction. These researchers determine that features other than surface errors account for the wide disparities in scores.

The NNS students in this study also score significantly lower with errors intact than the NS’s. In contrast, though, the NNS writers in this study do not score significantly lower than the NS group after error correction.

Thus, I conclude that for both groups of writers, the corrected scores are significantly better than the uncorrected scores. According to the statistical results, there is no significant main effect due to group. In other words, NS and NNS students in this study perform similarly.

Additionally, there is no group X score interaction. That is, the improvement from uncorrected to corrected scores does not vary for the NS and NNS papers.
Even though there is no significant difference between the NNS papers before and after correction, there is a significant variation between the papers before correction from the two classes: that is, the NNS essays scored significantly lower than the NS essays. After correction, though, no significant variation is seen between the scores for the two groups of papers.

**IMPLICATIONS AND SUGGESTIONS**

**THEORETICAL AND OTHER IMPLICATIONS**

The results of this study both support and contradict the findings from previous research. This effect alone is sufficient to justify the research done in this study. The reason is that such results indicate it is not true in all instances that NNS students are rated holistically for factors other than surface errors. Moreover, this study indicates the scores of NNS papers can improve comparably with NS papers after error correction.

Another major implication arising from this study is that NS raters who are not ESL-trained may, indeed, respond to ESL writing differently from NS writing. Thus, non-ESL trained English teachers often attend to error, even considerably, in responding to writing although they profess not to do so. This also implies that NNS students might fare
better and receive scores comparable to their efforts and abilities if they are taught by ESL-trained staff.

Yet another inference here is that holistic scores, even of NS student compositions, sometimes vary among raters, as evidenced by comparisons of ratings before and after correction on the same paper where scores actually decrease. The implication is that, to obtain reliability and validity, holistic raters require periodic retraining sessions, since their responses appear to vary.

Moreover, holistic rating sessions are supposed to rate partially by comparison with other papers read during a given sitting. Therefore, it seems reasonable to consider reading NNS papers together in a rating session separate from NS papers.

AREAS FOR FURTHER RESEARCH

Further research of the kind this study has conducted is appropriate. One reason for a continuation in this direction is the diversity of results in comparing NS and NNS writing scores before and after correction. Another reason is the dearth of studies currently existing in this area. One more suggestion for further research is to conduct studies similar to this one, but comparing scores awarded by ESL-trained and non-ESL-trained raters.
Additionally, a need exists for further research into the frequency of instructor response that varies from what is purported to be the chief writing criteria. Next, more studies are needed in other areas where present studies yield conflicting results. Some of these topics worthy of further investigation are response variations among instructors, agreement on a hierarchy of errors, harsh ratings because of perceived redundancy, negative response to error despite acceptable and understandable content, and response to variation in ability with syntax and rhetoric.

PEDAGOGICAL SUGGESTIONS

NNS students have differing cultural backgrounds and attitudes and unique problems with American English academic writing. This study suggests that this student population fares better in separate classes conducted by ESL-trained instructors. Further, Kroll (1990) suggests separate classes for NNS students for their individual levels of individual strength and weakness in rhetoric and syntax.

Additionally, writing instructors should ascertain that students can read, understand, and relate to their writing assignments. Readings should be discussed before writing assignments are given. Moreover, teachers might lead a class discussion concerning possible future writing prompts and how students might respond to them.
Next, I would recommend that great attention be given to choosing writing assignments and to writing specific prompts. Care should be taken not to assign culturally-insensitive, culturally-biased, or unfamiliar topics for students. For example, it would be insensitive to assign a topic that requests students to reveal more of themselves or their feelings than they wish. In addition, topics or prompts that assume superiority of American attitudes or activities are biased. Further, some students may not have had the opportunity for experiences that are taken for granted in American society, such as experiencing an exclusively American holiday. Moreover, the wording of the directions should be clear and the diction simple.

In many instances, NNS students respond well to collaborative efforts (Hvitfeldt 1986; Kantor & Rubin 1981, p. 77). Group writing assignments, at least some of the time or initially during a term, appear to be beneficial for some students. Moreover, a "workshop" classroom situation that focuses on ideas, various drafts of an assignment, and peer feedback seems to encourage some beginning English writers greatly.

Another area of concern is the time allotted for composing. It would be reasonable not to conduct timed writing assignments for graded assignments except as required for in-class examinations. The reason is that many
NNS students require more writing time than NS's (Kroll 1983, p. 139). Ungraded, practice timed writing assignments, however, continue to be of benefit to NNS writers.

Other suggestions relate to instructor response to ESL writing. For instance, teachers should suspend their judgment of an essay until it has been read entirely, perhaps even a second time (Land & Whitley 1989, p. 290). The focus of an instructor's teaching goal should be on the content, or ideas presented, in composing rather than on the form. Additionally, teachers should become as familiar as possible with alternative patterns of coherence and rhetoric.

NNS writers should not be penalized for experimenting and taking chances with writing in English; therefore, teachers should recognize and reward venturesome endeavors and growth. Nor should the focus of response be upon sentence-level surface mistakes.

As with any small, single piece of research, the results of this study should be interpreted with caution. Another reason to be cautious is that a margin of error still does exist.

While the results of this study vary considerably in some respects from those of similar studies, they do not in other respects. For instance, NS compositions are scored higher than NNS compositions before correction as well as
after correction, indicating that scores for NNS writers are influenced by minor surface errors, as other studies indicate (Zamel 1985, Fein 1980, Homburg 1984, McGirt 1984, Perkins 1990, Sweedler-Brown 1993a). Further, the results of this study tend to agree with Evola et al (1980). That is, the scoring for overall writing ability of these students does not appear to be affected by global errors such as coherence and cohesion.

In general, then, this study has contributed to the research on response to ESL academic writing. It has added to the few studies to date concerning instructor response to NNS surface errors, particularly in comparison with NS writing. It is hoped that further studies will be conducted on similar issues in order to add to the body of knowledge concerning the response to and rating of ESL writing.
APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX A</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biodata Form</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDIX B</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Prompt</td>
<td>79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDIX C</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Department Rating Scale</td>
<td>80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDIX D</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Essays From the Data Base</td>
<td>81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDIX E</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error List</td>
<td>93</td>
</tr>
</tbody>
</table>
APPENDIX A

BIODATA FORM

Questionnaire for Composition Students
CSUSB Department of English

Course: __________________________
Instructor: _______________________

1. Name _________________________

2. Address _______________________

3. Telephone (Day) _________ (Evening) _________

4. Best time you can be reached: _______________________

5. Native Language __________ Do you speak another
language at home? ______ Language: __________
Which language do you speak most frequently? __________
Which language do you usually speak with your friends?
English ______ Other ________________
If you speak another language with your friends, what is
the percentage of time you do so? ________%
Which language do you feel most comfortable speaking?
__________________________

6. Where were you born? ________________

If you were not born in the U.S., how long have you been
here? _______________________

Did you attend high school here? _______/_______ yrs.

Name and location of high school: _______________________

Visa status (if citizen, please put "U.S.") ______________

77
7. Which areas of writing do you have the most difficulty with? (Number 1-4 according to difficulty, "1" being the most difficult)

Grammar ___________ Organization ___________

Content (Ideas) _______ Spelling & Punctuation _______

8. What was your approximate GPA in your high school English classes?

Current (approximate) CSUSB GPA: ____________

SAT Verbal Score: _______ TOEFL/TWR Score: _______

EPT Score: ____________

9. List all CSUSB composition or writing classes you have taken (and grades earned):

10. List any ESL classes you have taken here, in high school, or elsewhere:
ENGLISH 101 COMMON IN-CLASS ESSAY

You will have 70 minutes to plan, write and review a well-organized essay on the topic below. Be sure to think and plan before you write, and spend some time at the end to proofread what you have written.

TOPIC:

In "Friends as Family," Karen Lindsay describes several different kinds of families. What sorts of "families" does Lindsay have in mind? What are the conditions for calling a group of people a "family?" It is possible to establish a family through non-genetic means today; for example, adopted children are part of their adoptive families, and gay men and women form lifelong relationships which they consider family. Some non-related groups of people even enjoy communal living and consider themselves "family." What conditions do these cases have in common that allow them to become families?
Papers should be scored for their overall quality, and the student should be rewarded for what is done well.

POSSIBLE SCORES (SIX-POINT SCALE AND SINGLE-CONCEPT SCORING MARKERS)

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPERIOR</td>
<td>6</td>
</tr>
<tr>
<td>COMPETENT</td>
<td>5</td>
</tr>
<tr>
<td>ADEQUATE</td>
<td>4</td>
</tr>
<tr>
<td>INADEQUATE</td>
<td>3</td>
</tr>
<tr>
<td>INCOMPETENT</td>
<td>2</td>
</tr>
<tr>
<td>INFERIOR</td>
<td>1</td>
</tr>
</tbody>
</table>
APPENDIX D

SAMPLE ESSAYS FROM THE DATA BASE

Persons interested in obtaining the entire corpus of compositions should contact the chairperson for this study:

Dr. Wendy Smith
English Department
CSUSB
Friends as Family

Lindsey describes three sorts of families; the nuclear family, which carries the blood bondage in them. For example the father, mother, children, and any other relatives. Secondly, the biological family, who are very close friends of the family, who we consider as relatives & treat them as relatives. Thirdly, the chosen family, a friend who you chose as a relative, someone that you love very dearly. the conditions for calling a group of people a "family" are simple. Its usually people who you trust and give much attention to as if they were part of you. When you call someone your family is usually someone very close to you and your family. Someone you grew up with. The only substance that would actually come between someone is not having the same blood running through their veins.

Yes, I honestly think that it is possible to establish a family through non genetic means. Nowadays, there is a lot of teenage pregnant women who willingly give their child for adoption. When on the other hand there is families or
husband and wife, who can’t bear any children. If the adopted parents get a newborn child they get it for a reason and are willingly to give them and show them all their love. Therefore, as the child grows older, and is raised with all that tender love. He will not care if has been adopted. Parents are also those who teach them their first steps, their first baby words, who shows them respect and makes them feel wanted and loved. Parents aren’t just those who give them birth, and just leave them like a piece of trash, those aren’t parents at all. In order for a mother or father to call a child his son or daughter he or she would have to gain that!!

The conditions that non-related groups of people called themselves families are usually people who work together, play sports together because through time you spend a lot of time together. Therefore, you get to know each other personally. When something is wrong or you have a personal problem you usually would go to them. Another thing would be doing things together, for example having get togethers, going to parties, movies, calling each other on the phone, etc.

I do agree with Lindsey that friends could be a family. For example, the biological family who is anybody outside the family, like co-workers. At your job you end up getting so close to them that even when you decide to leave. that
job, you always end up going back to visit them or calling
them and still be able to talk to them as if they were a
part of you. It is very wonderful to have someone outside
the family who you can trust and cry with and let all your
problems out. Mainly that person is the person you chose as
your Best Friend.
In "Friends as Family," Karen Lindsey describes the biological and the chosen families. In Lindsey's perspective, she did not believe in the "myth of the nuclear family." In fact, she doubted the fact that that kind of family ever existed. In the process of doing so, she disproved Lasch's theory of new narcissism. She gave evidence and proof that the nuclear family "... is more than a myth. It's also a lie" (399). For example, she pointed out that the nuclear family is closely related to the patriarchal family structure, where "The Father's Know Best."

Another kind of family that Karen had in mind is the "chosen family." She believed that it is possible to choose a group of people who could be as close to the self as anybody in a biological family could be. For example, she pointed out that in many cases, a chosen family can develop into a closer relationship.

The conditions that Lindsey established for calling a group of people a family are the following: in the biological family, the condition is to be blood related; for
example, any kind of relationship wherein there is a genetic relationship.

In the other type of family, the "chosen family," the conditions are to have a feeling of belonging or acceptance.

According to Lindsey, it's possible to establish a family through nongenetic means. First of all, the writer does not believe in the myth of the nuclear family, she said. From Lindsey's perspective, the so-called nuclear family was a patriarchal tool to keep women powerless and useless.

In addition to Lindsey's feminist point of view, people can choose their preferred families. The conditions are that women feel comfortable and free to do so.

The conditions that are necessary to allow "chosen" families to be families are that the members in any communal living arrangement can have a voice and make their own decisions.

In summary, Lindsey disapproves of the fact that the nuclear family used to be a pleasant time for women. In fact, she believed that the only reason that it appeared to work out is because women had no other choice, and that complaining about the patriarchal family structure was useless.
Families are the basis of everyone’s life, whether that family be blood related or adoptive. In every child’s life having someone to give you support and the attention you deserve shows you how to grow and love others. In Lindsey’s essay she says "The traditional family isn’t working," but I believe it is. Lindsey relates back in time to the era of "Mommy and Daddy & the kids." She reminds us of all the old television shows, for example--Father knows best and I Love Lucy. I believe that just as in the 60’s divorce was on peoples minds, but they had other, more important things to worry about. The War, education etc. People were getting married and having children all over the world. It was an era of learning and experimentation. Our world was growing immensely and we were on top of the world. I believe that a family consist of people who care and love you, and I do believe that the "traditional family is working. Whether you have a mother and/or a father you have someone who loves you. If you relate back to the t.v. show, little house on the prairie, the Ingalls had a Mom & Dad but they also brought in a son. They adopted, and their family grew.
I believe that it is very possible to establish a family through non-genetic means. It is what a single person believes a traditional family. In the Brady Bunch, they created a world famous family. Not genetically related they still loved and supported one another.

The major part of a family is not whether you are blood related but whether one can grow emotionally and intellectually in the right directions. Having the source of love acknowledgment, support and caring can give a child hope.

In some ways I do believe with Lindsey, as she describes as a myth. I think that family has grown in time with that actual title, as family because people agreed to it. In the earlier centuries, men ruled over women, and I believe that this is why it has stuck with us. Men (in the past) have seemed to be more powerful than women, allowing them to choose who they want, when they want them and how. Men would be the head of the household because they were "powerful" but I believe it is a person who is powerful or strong not the gender.

Lindsey also talks about wife beating and child beating, I believe that this occurs because of the economic situation and the immaturity of the parents. In a lot of cases the battered victim is that of a low-income family who are very young. In our society, children are having children. It
also could relate to cultures. In earlier centuries, it was natural for the husband to be non-respectful of his mate.

In our society today we can choose our family. Whether it be one husband or our children. I believe that non-communal living styles are up to an individual to decide whether or not they like it. They also have emotions as do we. They can love and respect others too.

In my opinion a family is a group of people who can share feelings with each other. They can love, acknowledge, respect and care for everyone, whether the family be genetic or not it is not up to a society to decide whether your feelings & emotions are that of theirs. You have your own mind and self, decide for yourself. The traditional family is up to your decision, no one can change that for you, that is why we live in America.
Midterm Question

Throughout this entire essay, Lindsey mentions a lot of family situations. She talks about the nuclear family, the biological family, and the families during the Golden Age, the Renaissance, and the Middle Ages. Also, she talks about the chosen family.

Lindsey seems to stress the nuclear family a lot. She first states that it is one of the "mythical concepts." Heaven and hell, the nuclear family, and the Russian Revolution are all what Lindsey says are myths. Lindsey says they are myths because "apart from whatever reality they have, the way in which we view them helps clarify, even shape, our vision of the world."

Lindsey wants us to return to the "good old days" of the nuclear family when Mommy stayed home and cooked, and Dad worked all day and spent the evening with his two kids. For us to believe this myth, she says that we women must turn into collaborators in our own oppression.
Lindsey believes that choice is a good one and that good things will come with that choice. In contemporary mythology, that choice guarantees happiness. Sometimes the chosen family mirrors the worst of biological families; for example, Charles Manson was the leader of a chosen family, as was Jim Jones.

The conditions for calling a group of people a "family" are that you have communication going. You are sharing ideas and feelings. Love is a quality which exists in a family. A quality of shared experience, history, and feelings can build a strong bond of love and affection within families. In a family, people can make deep and indissoluble connections with one another.

My personal definition of a family is that it has a feeling of acceptance and belonging. It exists when emotions are filled and dealt with.

Lindsey explains that it is not only spouses who have shared or do share each other's lives together that creates a strong bond, but it is also friends, neighbors, and co-workers who have equally strong bonds.

Lindsey gives examples of these nongenetic means, which television illustrates in a mythical way.

In the 50's, we had shows like *I Love Lucy*, *Ozzie and Harriet*, and *Make Room for Daddy*, which showed a mommy, daddy, and kids as the family.
In the 60’s, we had widows and widowers with kids becoming the next form of family. We had shows like The Partridge Family and The Brady Bunch.

The workplace family was brought up in the 70’s with shows like The Mary Tyler Moore Show and The Love Boat.

All of these families built that strong bond which kept all of them together. Communication, a shared history, and experience built a strong feeling of acceptance among them.

Overall, Lindsey has demonstrated family situations which she feels haven’t worked because of their being biological families. I think Lindsey is stressing the idea of the chosen family because you are able to work on having a strong bond to keep you together.
APPENDIX E

ERROR LIST

1. AGREEMENT
   Subject-Verb
   Determiner-Noun
   Pronoun-Antecedent

2. ARTICLE

3. PREPOSITION

4. SINGULAR vs. PLURAL NOUNS
   Count-Noncount
   Noun Adjuncts
   Object of Preposition

5. PRONOUN
   Case
   Reflexive

6. ADJECTIVAL INFLECTION

7. INTENSIFIER

8. VERB
   Form
   Passive
     "be"
     -en
   Auxiliary
     -en
     "be"
     -ing
   Particle
   Modal

9. PARTICIPIAL PHRASE
   -en
   -ing
   Auxiliary

10. GERUND vs. INFINITIVE

11. INFINITIVE "TO"

12. PREPOSITIONAL GERUND vs. VERB PHRASE

13. DERIVATIONAL MORPHEME
   Nominal
   Adjectival
14. **CONJUNCTION**
- Coordinate
- Correlative
- Subordinate
- Relative Clause Pronoun

15. **PUNCTUATION**
- Comma
- Possessive Marker
- Hyphen
- Quotation Mark
- Colon
- Exclamation Point
- Question Mark
- Underlining

16. **FRAGMENT**
17. **RUN-ON**
18. **COMMA SPLICE**
19. **SPELLING**
   - Misspelled
   - One Word vs. Two Words
   - Two Words vs. One Word
   - Homonym

20. **CAPITALIZATION**
21. **WORD ORDER**
22. **WORD CHOICE**
REFERENCES


Writing: Research Insights for the Classroom (pp. 191-210). New York: Cambridge UP.


