The effects of oral conferencing and written comments on the writing and revisions of ESL students

Ann Louise Kirk
Portland State University

1988

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds
Part of the Education Commons, and the English Language and Literature Commons

Recommended Citation

10.15760/etd.5688

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
AN ABSTRACT OF THE THESIS OF Ann Louise Kirk for the
Master of Arts in English: TESOL presented May 19, 1988.

Title: The Effects of Oral Conferencing and Written Comments on
the Writing and Revisions of ESL Students.

APPROVED BY THE MEMBERS OF THE THESIS COMMITTEE:

Marjorie S. Terdal, Chair
Jeanette S. DeCarrico
Shelley C. Rose
Jonathan O. Pease

This study looked at the effect of written and oral comments on
students' writing. The research hypotheses were that the use of oral
comments would improve the overall quality of the students' papers,
increase the length more, and cause more changes in content than the
use of written comments. On the other hand, the use of written
comments would cause a greater decrease in grammatical errors in the
students' papers than oral comments. The tests used to evaluate
these hypotheses were the holistic writing scale used by the Test of Written English (TWE), a word count, a content percentage scale developed by the researcher, and an adaptation of Brodkey and Young's Composition Correctness Score (1981).

The procedures used in the study were as follows: Three different classes were used. Each class was randomly divided in half. After the students had written the first draft of an essay assignment given by their teacher, they were assigned to either the written or the oral group. Students in the written comments group received written comments only on their papers. Students in the oral conference group had conferences with their teachers about their papers. Then the students wrote a second draft of their papers. At this point, the process was repeated; students wrote a second paper. After this first draft, those students who had received written comments on the first paper had oral conferences, and those who had had oral conferences received written comments. Again, the papers were revised based on these comments.

The results of the study gave qualified support of the hypotheses. It was the opinion of the researcher that this limited support was primarily caused by the different approaches and influences of the individual teachers, rather than a pure disproval of the hypotheses. In addition, the high variance in the scores and the limited subject pool further confounded results. Oral conferencing still has some demonstrated strengths which written comments lack, particularly in terms of the opportunity for feedback and the length of discourse, that merit its use in the evaluation.
of writing. The results certainly showed that it did no worse than written comments. Additional research needs to be done to isolate the effective evaluation strategies that teachers are using, and to demonstrate how to apply these strategies.
THE EFFECTS OF ORAL CONFERENCING AND WRITTEN COMMENTS
ON THE WRITING AND REVISIONS OF ESL STUDENTS

by
ANN LOUISE KIRK

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF ARTS
in
ENGLISH: TESOL

Portland State University
1988
TO THE OFFICE OF GRADUATE STUDIES:

The members of the committee approve the thesis of Ann Louise Kirk presented May 19, 1988.

Marjorie S. Terdal, Chair

Jeanette S. DeCarrico

Shelley C. Reece

Jonathan D. Pease

APPROVED:

John H. Cooper, Chair, Department of English

Bernard Ross, Vice Provost for Graduate Studies
I would like to acknowledge the following people for their support and help during the writing of this thesis: First and foremost, I would like to thank my advisor, Dr. Marjorie Terdal, for helping me to achieve the seemingly impossible. In addition, I wish to thank the other members of my committee, Dr. Jeanette DeCarrico, Dr. Shelley Reece, and Dr. Jonathan Pease for their guidance and advice. Lester Weber was extremely helpful with analyzing all of the data. I would also like to thank the students and teachers who were willing to participate in this study. Moreover, I wish to thank all the people who helped with the ratings in the study, John Sparks, Katrina Ratzlaff, Russell Davis, Tina English, Kim Henderson, and Becky Zweig. I also wish to thank my parents for their faith in me, and finally, my husband, Russell, for his unfailing support throughout this entire process.
# TABLE OF CONTENTS

| ACKNOWLEDGEMENTS | iii |
| LIST OF TABLES | vii |

## CHAPTER

### I INTRODUCTION
- Statement of the Problem | 1
- Background | 2
- Definition of Terms | 4
- Research Hypotheses | 6

### II REVIEW OF THE LITERATURE
- Teachers' Responses to Student Writing | 7
- Strategies to Aid in Revision and Assessment | 11
- The Writing Process in ESL | 16
- Expectations of Writing | 17
- Students' Reactions to Teachers' Comments | 18
- The Oral Conference | 20

### III METHODOLOGY
- Research Methodology | 24
- Subjects | 24
- Design | 26
- Procedures | 26

**Quality: TWE Rating Scale**
**Length: Word Count**
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV FINDINGS</td>
<td>36</td>
</tr>
<tr>
<td>Analytic Techniques</td>
<td>36</td>
</tr>
<tr>
<td>Description of Findings</td>
<td>37</td>
</tr>
</tbody>
</table>

  * Hypothesis One  
  * Hypothesis Two  
    * Changes in Content  
    * Changes in Length  
  * Hypothesis Three  

  * Other Findings | 46 |

V SUMMARY AND DISCUSSION | 51 |

  * Conclusions | 52 |
  * The Teacher Variable  
  * Other Confounding Variables  
  * Tests and Measures  

  * Implications | 64 |
  * Implications for Teaching  
  * Implications for Further Research  

BIBLIOGRAPHY | 68 |

AFFENDICES

A INFORMED CONSENT | 72 |
B TWE SCORING GUIDELINES | 73 |
C SAMPLE RATINGS: ERROR RATIO SCORE | 74 |
D LIST OF ABBREVIATIONS | 90 |
E TWE SCORES: BOX AND WHISKER PLOT | 91 |
F CHANGE IN CONTENT: HISTOGRAMS | 92 |
G CHANGE IN LENGTH: HISTOGRAMS | 94 |
H ERS AND TWE: SCATTER PLOTS | 96 |
APPENDICES

I  TRANSCRIPTS OF TEACHER/STUDENT CONFERENCES AND
    COPIES OF TEACHERS' WRITTEN COMMENTS. . . . . 101
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>25</td>
</tr>
<tr>
<td>II</td>
<td>27</td>
</tr>
<tr>
<td>III</td>
<td>30</td>
</tr>
<tr>
<td>IV</td>
<td>40</td>
</tr>
<tr>
<td>V</td>
<td>40</td>
</tr>
<tr>
<td>VI</td>
<td>42</td>
</tr>
<tr>
<td>VII</td>
<td>44</td>
</tr>
<tr>
<td>VIII</td>
<td>45</td>
</tr>
<tr>
<td>IX</td>
<td>46</td>
</tr>
<tr>
<td>X</td>
<td>48</td>
</tr>
</tbody>
</table>

**LIST OF TABLES**

**TABLE I**

Subject Assignment to Groups

**TABLE II**

Essay Assignments for Each Group

**TABLE III**

t-Tests for the Difference Between First and Second Drafts for Both Oral and Written Comments for the TWE Scale

**TABLE IV**

Means of the Differences of TWE Scores for Papers Receiving Oral and Written Comments Using the Two Sample t-Test for Equal Variance

**TABLE V**

Means of the Differences of TWE scores for Papers Receiving Oral and Written Comments Using the Paired t-Test

**TABLE VI**

Percentage of Content Change, Two Sample t-Test and Paired t-Test

**TABLE VII**

Change in Length: Differences Between Draft I and Draft II

**TABLE VIII**

t-Tests for the Difference Between Error Ratio Scores for the First and Second Drafts for Both Oral and Written Comments

**TABLE IX**

Change in Error Ratio Score: Differences Between Draft I and Draft II

**TABLE X**

Correlation Matrix for the TWE and ERS for W.I, W.II, O.I, and O.II
TABLE  

XI  Correlation Matrix for Wordcount with TWE and ERS for  W.I, W.II, O.I, and O.II ........................ 50
CHAPTER I

INTRODUCTION

STATEMENT OF THE PROBLEM

Current theories of writing perceive writing as a process, not just a final product (see Boiarsky, 1984, and Elbow, 1981). If this view is accepted, effective writing instruction should reflect this view. Teachers should not be seen as just graders evaluating final results, but should actively intervene in the process of writing itself, guiding students and helping students improve their ability to write and revise. Teachers' comments on finished copies of papers serve only to justify grades. They neither help the student to write better, nor aid the student in doing effective revisions (see Brannon and Knoblauch, 1982). The purpose of teaching writing is two-fold. First, it is to teach students to become writers, to be able to use written language clearly and effectively. Second, it is to teach them to be able to decide whether or not something they have written is clear and effective, and know the steps they should follow toward achieving this. This seems essential if teachers wish their students to carry something of what they have learned beyond the teacher's specific classroom (see Murray, 1982). It is entirely possible, through the process of revision, to have a teacher so take over a student's text, that though the final version may be error free, even
eloquent, the student may have no idea how to do it again, and learns basically to rely upon teachers' comments to improve writing (see Brannon and Knoblauch, 1982; Sommers, 1982; and Sullivan, 1986). Though this type of situation is common, it is not acceptable if teachers are to create truly effective writers, in either their first or second language. Effective writing is writing that not only is correct, but also gets across its intended meaning and hopefully even moves its audience.

With this in mind, I have chosen to look at two types of intervention that teachers can make during a student's writing process: oral comments in a conference situation, and written comments given on the student's paper. It is the opinion of the researcher that oral comments, by their nature of being two-way communication so that both the teacher and the student can clarify meaning within the conference, are more effective than written comments, which are one-way communication and do not permit students to clarify either the teacher's or their own intent in improving subsequent drafts of an essay.

BACKGROUND

In the last ten years there has been a plethora of literature written on the evaluation of writing. Much of it deals with the role of the teacher in the writing process of students. For example, Sommers (1982), looks at different types of teacher responses and their effect on first language writers. Zamel (1985) did a similar study with ESL students. Griffin (1982) and Brannon and Knoblauch
(1982) also have studied the role of the teacher in a student's writing. Many of the articles are testimonials, techniques that worked for specific teachers in specific situations, (see Marshall, 1986; Rose, 1982; and Urzua, 1987). Few studies have systematically evaluated different forms of teacher evaluation and the effects of these evaluations on student revision. This is particularly true of second language research, what research that has been done is primarily in the field of first language writing.

During the course of my teaching assignment as a lecturer in English as a Second Language, I have had the opportunity to teach a variety of writing classes, with varied amounts of success. As a novice teacher, I felt very much at a loss as to how to approach evaluation of student writing to best help students actually improve their writing skills. The literature in the field, particularly Elbow (1981), Miller (1982), Bioarsky (1984), Moran (1982), and others, offered good theoretical information on current concepts of the writing process, but little practical advice as to how a teacher could effectively bring those concepts into the classroom, and particularly how a teacher should go about adapting them for use with second language writers.

Based on this literature, and experiences in the classroom, I began to look more closely into the use of oral conferencing, as opposed to written comments, during the students' writing process (see Carnicelli, 1980; Simmons, 1979; Marshall, 1986; and others). This thesis is an attempt to study this matter more closely, in hopes of discovering to some degree the influence that oral conferencing
and written comments have on subsequent drafts of ESL student writing.

**DEFINITION OF TERMS**

There are several terms which must be defined as they appear in this research. ESL students, unless otherwise noted, refer to either international students or non-native residents at the university level enrolled in the Center for English as a Second Language (ESL), or the program of English for Non-native Residents (ENNR) at a university in Oregon. In the Center for ESL, the Michigan Test of English Language Proficiency (MTELP) is used to place students. A student with a MTELP score of 51 to 60 is considered Lower Intermediate, 61 to 70 is designated as Upper Intermediate, and 71 to 85 is Advanced. Students with a score of 85 or higher on the MTELP are generally released from the ESL program, although it is the Test of English as a Second Language (TOEFL) which is used by the University to determine whether or not a student is accepted. The ENNR program divides students as follows: A MTELP score of 66 to 75 is Upper Intermediate, and 76 to 85 is Advanced. Although the scores are not the same for both groups, they are similar enough to permit joint classification. Moreover, for this study, the level of the ESL students in the program is not as important as their development over two drafts of a single writing assignment.

Oral conferencing is defined in this study as a one-to-one conference between an individual student and the writing teacher.
discussing the first draft of a writing assignment and possible revisions to it. All oral conferences took place in the teachers’ offices and were scheduled by the student and teacher. They varied in length from ten minutes to half an hour, with most being about twenty minutes. Written comments are the notes and comments that the writing teacher writes on the student’s first draft, discussing this draft and possible revisions. Students who received written comments received no oral feedback on their writing during this phase of the research. The entire process was done twice, with students who received written comments on the first assignment receiving oral conferencing on the second assignment.

For the purpose of this study, three types of revisions that occurred from the first to the second draft of each paper were reviewed. These types were, 1; revisions in content, alterations in the meaning of the paper, 2; revisions in length of the paper, and 3; revisions in surface errors (not errors of organization or argument), for example changes in grammar, spelling and sentence structure. In addition, changes in overall quality of the work were looked at, as judged by the holistic scale used in the Test of Written English (TWE) developed by Educational Testing Services (ETS). This holistic rating is a six-point scale, with a paper receiving a six considered to be an extremely effective paper, in which the writer has fully addressed the task at hand, has had very few errors, and has written a paper that reads like a competent paper written by a native speaker. A paper receiving a one is one in which the writer has not been able to master the writing task assigned and the writing is not
comprehensible due to errors. The other ratings range between these two extremes.

**RESEARCH HYPOTHESES**

There are three basic hypotheses that this study tested.

1) ESL students' compositions which have received oral comments during the writing process will show greater improvement than those which have received written comments, as determined by the holistic rating scale used with the Test of Written English (TWE) part of the Test of English as a Foreign Language (TOEFL).

2) Compositions receiving oral comments will show more changes at a content level, and will increase in length more than compositions receiving written comments. The tests used to measure hypothesis 2 are a percentage based scale looking at the percentage of change in the content of the essay from one draft to the next and a simple word count.

3) Compositions receiving written comments will show a greater decrease in errors of grammar and usage than those receiving oral comments. The test used to measure hypothesis 3 is an error ratio count based on the Composition Correctness Score developed by Brodkey and Young (1981).
CHAPTER II

REVIEW OF THE LITERATURE

Current research in writing supports the view of writing as a process (see Elbow, 1981, Bowers, 1984, and others) instead of concern only with the product. With this change in approach to the theory of writing, from product-only to process, a change in theories of responding to student writing is also occurring. Comments made during the writing process are made to generate revision, not justify a grade. A great deal of research has been done recently into different ways of responding and evaluating student writing and their effects on students, teachers, and writing instruction.

TEACHERS' RESPONSES TO STUDENT WRITING

Sommers (1982) gives an overview of teachers' responses to student writing and ways of responding that best help the student. In her view, the goal of the writing teacher is to help the student to internalize a sense of reader and to offer assistance during the composition process. As such, the teacher must provide comments which give this assistance. Without any comments, Sommers feels, students will not revise. But too often the comments that the student receives shift her attention from "This is what I want to say," to "This is what you the teacher are asking me to do" (p. 150). "Teachers comments can take students' attention away from their own purposes in writing a particular text and focus that
attention on the teacher's purpose in commenting" (p. 149). When this occurs, the student becomes insecure, unwilling to venture much of herself, for fear of being contrary to the teacher's purposes.

If a teacher comments on errors in usage, style, and diction in the first draft, it causes the student to see the piece as completed, needing only surface changes. This results in a general unwillingness on the part of the student towards further revision. If there are flaws in content, argument or meaning, the student is resistant to changing them. Frequently writing teachers make comments on both surface errors and general structure and content. These comments are contradictory and confusing, for if the student revises the content, then often the offending sentences disappear or right themselves, and if the student fixes all the surface troubles, she will most likely not want to go back and throw them out altogether.

Another basic problem that Sommers targets in teachers' styles of commenting is the comments themselves, (see also Siegel, 1982, for results of a study on the types of responses new composition faculty give to student writing). For the teachers she studied, the comments were generally too vague to help the students successfully revise and did not seem to be significantly related to the actual text. "Most teacher's comments are not text-specific and could be interchanged, rubber-stamped, from text to text" (p. 152). But she noticed that there was an

...overwhelming similarity in the generalities and abstract commands given to students. There seems to be among teachers an accepted, albeit unwritten canon for commenting on student texts (p. 153).
The vagueness of these comments causes even more difficulty because of the fact that they are seldom presented with any relevant strategies to deal with them.

Vivian Zamel (1985) conducted a similar study with ESL teachers and found very similar results. Far too much focus is placed on grammar and surface organization, and comments are generally vague and useless. One aspect that seems peculiar to ESL, however, is that teachers often misinterpret intended meanings. They correct students' work, that is, tell students exactly how to rewrite a passage, and in the process, change the students' meaning. Most students assume that the teacher is right, give up the authority of their texts and rewrite them in the teacher's image. This type of misinterpretation is particularly evident in ESL because of the difficulty students may be having with mastery of English grammar and sentence structure. Meanings are often obscured because of incomplete control of grammatical structures.

In order to gain some insight as to how to change this dismal situation, C.W. Griffin (1982) studied current theories of responding to student writing and outlined three basic criteria he feels critical to these emerging theories: "...our orientations, our verbal responses, and our students' reactions to our responses" (p. 296). Orientations are the different ways in which composition teachers approach writing and responding to student papers. As the above studies would indicate, generally the focus has been on structure over content, but Griffin also cites studies which have gotten opposite results. Additionally, different teachers evaluate
writing differently; the same teacher may even do so at a different
time. One should keep in mind all the variables of orientation and
expectation that a teacher brings to the task of reading student
writing. In any teacher's verbal responses, there is a variety of
types of responses that can be made, looking at different parts of
the students' writing.

The major question confronting any theory of responding to
student writing is where we should focus our attention.
And when we do focus in a certain place, what assumptions
are we making about the nature and function of writing?
(p. 299).

Teachers may look at the content or the structure of an essay, may
question a student or may correct her. These responses reflect the
different attitudes towards writing and response that exist in the
field today. Ultimately teachers must look at how students respond
to their comments to know the comments' effectiveness. To do this,
teachers must have some understanding of the process of revision and
the effect of different comments on this process. Research suggests
that "In essence, experienced writers view revision as part of the
writing process; inexperienced writers view revision as the tail end
of a process already completed" (p. 301).

To overcome this view of revision is the responsibility of the
teacher. Teachers should approach student writing as process, not
edit it as if the text were autonomous (Moran, 1982) This belief
comes from a theory of reader-reponse based criticism as opposed to
text-based criticism. In text-based criticism, the text is seen as
the holder of reality (p.4). "There is an ideal form; this form is
placed like a template over the existing essay, and the teacher notes
where the essay does not fit" (p. 7). Teachers should instead respond to student writing as real readers, not teachers, bringing to the task all their experiences as natural readers, accepting the variance in evaluation and interpretation that this necessarily brings.

The writer is a human being, and so must consider the human condition. The writer is free, autonomous, and must be so. The writer is also constrained by the context into which the writing will fall. In school-writing, that context is a teacher who is a reader responding to a text. Different teachers will respond in different ways (p. 16).

It is the responsibility of the teacher-reader to read naturally, not hunt for errors and to accept the natural variance in response.

Robertson (1986) and Heffernan (1983) also discuss the importance of reading student writing, not for errors, but for what lies behind the errors. Robertson points out that in normal conversation, one would hardly stop a speaker telling an exciting story to tell him that he drops his final -d's in the past tense, but teachers feel no compunction about doing this to student writing. Heffernan discusses another pitfall of error-focused responses.

To say anything to students in writing is to realize anew the difficulty of saying anything at all without making what could be called an error, without breaking one of the scores of rules that we insist our students know and obey (p. 4).

Both think that a teacher should read through error, not for error.

STRATEGIES TO AID IN REVISION AND ASSESSMENT

Brannon and Knoblauch (1982) support Griffin's view of the inexperienced writer. They think that this vision of revision as editing is partly due to the types of writing evaluation that most
students receive on most of their writing. As Sommers pointed out, most teacher responses are in the form of corrections of surface errors. To break away from these types of responses and the resulting revisions, Brannon and Knoblauch suggest that the teacher and student-writer work together to achieve union between intention and effect. This is best achieved if both answer general questions such as: "What did the writer intend to do? What has the writer actually said? How has the writing done what it is supposed to do?" (p. 162). Moreover, the teacher must work to break away from comparing student writing to some sort of "Ideal" text and must give back to the student the same rights given to professional writers, authority and control over writing. They feel this can best be achieved in the oral writing conference.

Several different scales and measures have been developed to aid teachers and students in evaluating writing. Jolly (1985) found that the majority of teachers in her study did not know how to evaluate student writing. When they lack this knowledge they often become editors (p. 4). She trains teachers to look at four basic (and familiar) areas, content, organization, grammar, and mechanics (p. 5). She recommends that the first two be dealt with in the first revision and the last two in a later revision. This is an attempt to help teachers see writing as a whole unit, not just as errors that need correction.

Elbow (1981) has developed two different categories for response to writing: criterion-based response and reader-based response. Criterion-based response evaluates the text as it compares to a
specific set of criteria. These criteria can be different depending upon the situation and the needs of the writer. Reader-based response is the personal reaction that the reader has to a piece of writing, not comparing it against another text or set of criteria, but just giving his or her personal response to the text. Elbow has created lengthy lists of questions that can be used to facilitate both types of responses. Pedersen (1984), Marshall (1985), and Anandam (1983) have all developed different computerized checklists and evaluation forms that teachers can use to comment on and evaluate student writing.

Miller (1982) looked at the criteria that professional writers use to evaluate their own writing and compared these criteria to student writers. She found that student writers use only one of the three evaluation criteria professional writers use. The only criterion student writers used was whether or not their audience, most likely their teacher, liked their writing. Professional writers also looked at "...whether the piece fulfills the writer's intention, and whether the writer has learned from the writing" (p. 179).

Lynch (1982) and Beach and Eaton (1984) have all developed self-assessment scales for students to use in evaluation of their own writing. Both studies indicate that guided self-assessment is critical in training students to evaluate their own writing. Students lack the skills to even begin this process without help and need assistance from the teacher in learning how to do self-assessment. Lynch proposes questions to help direct students throughout the writing process, at the pre-writing, rough draft,
revision, and final draft stages (p. 310). Beach and Eaton's questions are more general, and are used after the rough draft is finished, to help the student begin to understand how her writing works or fails to work.

In a later article, Beach (1986) outlines techniques that teachers can use in assessing student writing in a conference. The ultimate goal is to teach these techniques to students through different methods, including modeling, so that they can use them themselves. He has developed a model of assessment to facilitate this training. The three major stages of the assessment process are describing, judging, and selecting/testing out revisions (p. 57). Within each stage he provides guided assessment forms. A teacher analyzes the areas of assessment a student is having difficulty with and then uses various techniques to improve the student's ability to assess in this area.

Raymond (1982) discusses criteria for judging methods of writing evaluation of first language writers. Since the whole concept of language is not easily quantifiable, and is different for each language user, he thinks that all systems of evaluation should be carefully scrutinized. He sets up several criteria for judging them, for example using different kinds of writing tasks on different days, and having more than one rater.

Perkins (1983) conducted a study reviewing tests and measures used to evaluate the writing of ESL students. Though his analysis indicates that holistic and analytical scoring schemes are the most effective evaluation methods for ESL writing currently, he cautions
that "...no test or scoring procedure is suitable for all purposes" (p. 666). All judgement of writing is based on values, which are difficult to quantify and are not the same for everyone. Tests to evaluate student writing should be studied closely, to see if they are consistent with and appropriate for the purposes for which they are intended.

Brodkey and Young (1981) have developed a "Composition Correctness Score" (p. 159). This score is based upon the number and severity of errors in a student's writing. Although it seems to have low content validity, since it looks only at the errors in the writing, and not at the writing itself, Brodkey and Young have achieved high reliability in using it as a placement tool in their ESL program.

Homburg (1984) studied the objective validation of holistic ratings of writing tasks. He looked at various objective measures, such as T-units, measures of error and measures of composition length as they compared to holistic ratings given in the writing portion of the Michigan Test of English Language Proficiency (MTELF) and found high correlation between these methods. His findings suggest a "funnel model" (p. 102) that can be used when making holistic evaluations, moving from broad categorization to narrow categorization before the final evaluation is made. Greenburg (1986) discusses evaluation of the TOEFL writing test (TWE), and the holistic scoring procedures used in it. Different types of writing tasks are used in different TWE tests; including interpretation of a graph, supporting a position, or comparison and contrast. Her
findings indicate that though the reliability between the different forms used in the TWE is not great, the test itself is one of the soundest tests of writing ability she has encountered.

THE WRITING PROCESS IN ESL

Urzua (1987) and Zamel (1983) have both conducted case studies of the development of ESL writers. Urzua worked with children and Zamel worked with adults in a college ESL program. In Zamel’s study, the process skilled writers used in writing was compared to the process an unskilled writer used. She found that the skilled writers made more revisions and their revisions were more substantive than the unskilled writer. The unskilled writer’s revisions were only at the surface level. She viewed her writing as finished after the first draft, and saw future revisions as only to fix up her first attempt. However, her writing was of much poorer quality than the other writers. The other writers cycled through writing, evaluation and revision throughout their writing process. This is in line with first language research conducted by Miller (1982) and Brannon and Knoblauch (1982).

Urzua’s study found that even children are capable of finding their "voice" and writing effectively with an appropriate sense of audience when they are given feedback, both by the teacher and by fellow students during the writing process. This research supports the findings of research done in first language writing (p. 295).

This research, and other studies replicating first language studies with ESL students, (see also Marshall, 1986; and Robb, 1986),
support the position that the process by which ESL students write, and the effect of teachers' comments on this process, is not significantly different from the process of first language writers.

EXPECTATIONS OF WRITING

If teachers are to prepare students to write in academic situations, it seems primary that they find out what kinds of writing students are expected to do once they leave their composition classes. Weiser (1981) conducted a study to gain just this information. He sent a questionnaire to faculty targeting the types of writing students do in their courses and their perception of the errors and quality of student writing. In most introductory level courses short answer exam questions are the only writing students have to do. As the classes become more advanced, more writing is required. Few assign a percentage of the grade on the basis of correctness. The English department was more critical of grammatical and stylistic errors than any other department. Mallonese and Breihan (1985) did a study similar to Weiser's, instead focusing on teachers' responses to student writing across disciplines. They found that there was little consensus across disciplines in any area except the feeling that students were not writing as teachers wanted. The researchers developed a set of criteria to help teachers in all the disciplines gain consensus in the types of comments used on student writing.

Boiarsky (1984) conducted a questionnaire looking at the way writing authorities view the teaching of writing. Her results show
that the experts favor writing as a process approach. The authorities targeted a number of objectives that a teacher should use in the writing classroom; for example, students should select their own topics slightly more than half the time, and slightly more than half the class time should be spent actually writing.

STUDENTS' REACTIONS TO TEACHERS' COMMENTS

Burkland and Grimm, (1984), Ziv (1982 and 1984), Faigley and Witte (1984), Onore (1984), Semke (1984), and Sullivan (1986) all discuss the students' responses to teachers comments. In general they have found that students find most teachers' comments extremely difficult to understand and follow. In fact, the findings of Burkland and Grimm indicate that students frequently simply cannot even read their teachers' comments. To be effective and promote revision, teachers' comments must be text-specific, not vague generalizations such as "awkward" or "unclear" (Onore, 1984:3).

If teacher comments are to be helpful, then they should not only indicate the problems a reader is having with a text but should also suggest strategies which would help the student writer solve these problems. (Ziv, 1982:5-6)

Although students say they prefer to receive grades, Burkland and Grimm found that students pay more attention to comments on final papers and apply them more to future papers if a grade is not given. Additionally, most students want more comments from their teachers on their papers, even on final drafts not to be revised.

Ziv (1984) has created categories for teachers' comments and has analyzed the effect of these different comments on students' revisions. Her categories are: explicit comments, in which the
teacher points out a problem or issue and gives a specific solution; implicit comments, in which the teacher points out a problem or issue and may or may not suggest revisions; and corrections, in which the teacher rewrites the problem area, correcting any perceived errors. She found that students benefitted most from explicit comments, which are text-specific, and least from teacher corrections, which appear to the students as teachers appropriating their texts (pp. 372-375).

Berkenkotter (1984) and Onore (1984) (see above), also discuss the different approaches to revision that students take, regardless of teacher or peer comments. Berkenkotter, whose study was conducted in the context of peer response groups, describes three types of revisers: the resistant reviser, who will make only very surface changes and sees the first draft as complete, not needing revision; the inner-directed reviser, who incorporates comments in revision, but also has a strong sense of authority over the text; and the reviser with a crisis of authority, who tries to incorporate and accommodate to all comments in revision, regardless of their impact on the text. Of these three, Berkenkotter feels that the second type, the inner-directed reviser is the most successful reviser.

Other studies looking at the effectiveness of peer response groups are Grimm (1986), Benesch (1984), Benesch (1985), and Flynn (1982). These studies also looked at different ways to improve the effectiveness of these peer response. Benesch (1984) thinks that effective peer response relies on permitting students to use their own language, or way of speaking, and the relevance of their own
experiences in the peer response, rather than the teacher's code.

...while students do have a lot to learn about giving and receiving effective feedback -- as we all do -- they have had previous experience as writers, talkers, and collaborators which can serve as the point of departure for learning to respond to their peers' writing (1984:1).

To achieve successful peer response, the teacher intervenes in two primary ways: 1) on-going modeling of effective feedback, and 2) guiding discussions about what is happening in the groups and what changes may be desirable in the kinds of feedback used. These techniques combat what Benesch feels to be the greatest problem in peer response, "a lack of true exchange" (p.8).

THE ORAL CONFERENCE

The use and effectiveness of oral conferencing is discussed by Johnston (1982), Simmons (1979), Marshall (1986), Carnicelli (1980), Rose (1962), and Murray (1982). Johnston discusses the importance of giving responses to student writing that are not judgemental and lead to self-assessment by the student writer. This can be done through questioning the student, and giving personal and empathetic responses to the student's writing.

Simmons' study is a pilot study of a program using Roger Garrison's one-to-one method of teaching composition. The Garrison model uses short, in-class conferences with each student while the other students are writing or working with tutors. The student rewrites a paper several times and each time a different aspect is looked at in the conference. The students (particularly remedial
ones) given this treatment surpassed their counterparts in traditional writing classrooms on holistically graded composition exams. In addition, student and teacher morale was higher for the treatment groups and attrition rate was lower.

Marshall uses an adaptation of Garrison's method in an ESL classroom. Her underlying beliefs for use of the conference method are that "Learning how to approach a writing task is more important than producing a final copy" (p. 4). "The role of the teacher is to be a guide...As a mentor, the teacher's role is to provide the student with strategies for approaching a writing task (p. 6). The most effective method for achieving these goals, and the goal of making students into independent writers is through the conference.

Because writing is a communication process, the writer needs a response from others in order to know whether his effort to communicate has succeeded or not...After feedback, writers, especially student writers, are much more likely to be able to re-see, and thus revise, their written effort. The revisions produced from such feedback are not the surface corrections of local errors (which do little to improve communication), but instead, the rethinking and rewriting of the thought so as to clarify the piece as a whole (pp. 7-8).

During conferences, the student and teacher work together to decide what changes should be made. Marshall has had far greater success with her ESL student writers using the conference method than she had using written comments. The format for her conferences was based on Garrison, Simmons (above) and Carnicelli (1980).

Carnicelli defines the conference method he uses as based on the view of writing as a process, with three broad stages, "prewriting, writing, and rewriting" (p. 102). He outlines a list of priorities to consider when writing and when reading or conferencing on student
writing. These are: "content (ideas and information), point of view (purpose, persona, audience), organization, style (diction and syntax), and mechanics (grammar and punctuation)" (p. 103). These must be dealt with in this order in the writing conference.

He defines six tasks that the teacher should use in a conference:

The teacher should read the paper carefully... The teacher should offer encouragement... The teacher should ask the right questions... The teacher should make specific suggestions for revising the paper... The teacher should evaluate the paper... [and] the teacher must listen to the student (pp. 111-117).

It may appear that Carnicelli differs from many supporters of the conference method, (Johnston, Marshall, Brannon and Knoblauch and others) in that he advocates use of teacher evaluation in the conference process. However the type of evaluation he is referring to is subjective evaluation, the teacher's personal reaction to the writing.

The tasks Carnicelli outlines are straightforward and based on sound underlying principles. Unfortunately conferences are not as simple as following six easy steps. Carnicelli offers suggestions for applications to relieve this difficulty as well as provides examples of a successful and an unsuccessful conference and why they are such.

Rose (1982) supports the conference method of responding to students because it is more effective in targeting breakdowns in meaning and effectiveness in writing. When something is unclear in a student's writing, the teacher has the opportunity to ask the student what the intended meaning was, rather than just scrawling "unclear"
in the margin. By using the conference, the student, in turn, has the opportunity to clarify her intended meaning and the concrete experience that what she had intended to say is not what her audience, the teacher, understood.

Another supporter of the oral conference is Murray (1982). He also feels that the oral conference is the most effective way of returning authority of a text to the student. The oral conference is also effective in helping the students learn to write and revise for themselves, seeing themselves as their first reader.

The literature in the field supports the view of writing as a process. The role of the teacher should be as a facilitator of this process. Teachers should avoid taking control of the writing process away from the student and should provide comments that promote the process of writing. Research done in the field of ESL writing indicates that the process of ESL writers is very similar to the process of native English writers. The effectiveness of the oral conference in achieving better control and therefore better writing within this process is also indicated by the literature.
CHAPTER III

METHODOLOGY

RESEARCH METHODOLOGY

The methodological design used in this study was quasi-experimental. It could not be truly experimental because of the inability to randomize the subject group. It consisted of a pre-test (the first draft of each essay), a treatment (either oral or written comments), and a post-test (the second draft of each essay). Different variables were evaluated before and after the treatment. All subjects received both treatments in the course of two assignments.

SUBJECTS

The subject population for this study were students of English as a Second Language, enrolled in either an English as a Second Language (ESL) program or an English for Non-Native Residents (ENNR) program. They were at the lower-intermediate, upper-intermediate, or advanced level. Operationalization of these levels can be found in Chapter I, in the section defining terms. The subjects were assigned to three different teachers based upon the program they were eligible for and their level of ability. The students' assignment to the teachers was not random because of the limitations in size of the program and a desire not to interfere with the natural course of the
instruction for which the students had paid. The treatment was conducted twice, with all students receiving both treatments before the end of the study. Within each class, however, the students were randomly assigned to either the oral conferencing group or the group receiving written comments on the initial assignment. These groups were reversed for the second paper. At the end of the study, the subject pool contained 29 students total from the three classes. The first group, the advanced group, contained the largest number of students; 16. The second group, the intermediate group contained 9 students; and the third group, which was the lower-intermediate group, contained 4 students. Table I shows the breakdown of students per class.

**Table I**

SUBJECT ASSIGNMENT TO GROUPS

<table>
<thead>
<tr>
<th>Group</th>
<th>Teacher's Number</th>
<th>Number of Students</th>
<th>Level of Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>16</td>
<td>Advanced</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>9</td>
<td>Upper-Intermediate</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>4</td>
<td>Lower-Intermediate</td>
</tr>
</tbody>
</table>

The initial subject pool had been 43, but for various reasons the attrition rate for this study was rather high. The study was part of the regular instruction provided for the students, so there was no reliance upon students who chose to volunteer, except that all students had to sign a written consent form. Only 5 students declined to participate initially. A copy of the consent form that all students and teachers involved signed can be found in Appendix A.
The anonymity of all students was strictly regarded, in compliance with the federal regulations regarding the use of human subjects.

DESIGN

The independent variables were the two treatments, either written or oral comments given by the teacher after the first draft and before the second draft. The dependent variable was improvement from draft one to draft two. Improvement was measured through the use of a holistic scale based upon the scale used in the Test of Written English (TWE), developed by Educational Testing Service (ETS). Additionally, the dependent measures include a scale used to determine the ratio of errors in each essay, a percentage scale to determine changes in content, and a word count to determine changes in length.

PROCEDURES

The experiment was conducted as follows: In each class, students were assigned either to a group receiving oral comments first or to a group receiving written comments first. In the course of the study, all students received both oral and written comments, either on their first assignment or on their second. Each student wrote the first draft of an essay assigned by their teacher. The topics for the first assignment varied with the class. One group did a comparison/contrast paper, one group a description of a place, and the third group wrote an assignment telling "how to" do something (for instance, how to succeed in college), a description of a
process. After the first drafts were completed, clean, unmarked copies were made for the ratings used in the study. Then the students received the treatment, either comments written on their essays, or an oral conference with their teacher. Copies were made of all the written comments and all conferences were audiotaped. The students then revised and rewrote their essays. After they turned in the second draft, clean copies were again made for the purposes of the study.

At this point, the entire study was repeated again, with students who had received written comments the first time receiving oral comments the second time and vice versa. As with the first group, the second assignments differed for each group. One group wrote persuasive argument papers, another descriptions of a personal experience that promoted growth, and the third a comparison/contrast of books, television shows, or movies. Table II outlines the different assignments given to each group. Once the data were collected, rating was begun. Four different rating scales were used in this study.

### TABLE II

**ESSAY ASSIGNMENTS FOR EACH GROUP**

<table>
<thead>
<tr>
<th>Group</th>
<th>Paper Number</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>I</td>
<td>Comparison/Contrast</td>
</tr>
<tr>
<td>A</td>
<td>II</td>
<td>Persuasive Argument</td>
</tr>
<tr>
<td>B</td>
<td>I</td>
<td>Description of a Process</td>
</tr>
<tr>
<td>B</td>
<td>II</td>
<td>Comparison/Contrast</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>Description of a Place</td>
</tr>
<tr>
<td>C</td>
<td>II</td>
<td>Description of an Experience</td>
</tr>
</tbody>
</table>
To determine changes in overall quality, the rating scale used in the Test of Written English (TWE) was used. To determine changes in length, a word count was used. Changes in errors were determined using a ratio scale based upon one developed by Brodkey and Young (1981). Finally, changes in content were determined by using a percentage scale developed by the researcher.

Quality: TWE Rating Scale

The scale used to determine the overall quality of each paper is a scale developed by the Educational Testing Services (ETS) for their Test of Written English (TWE), which is used as a part of the Test of English as a Foreign Language (TOEFL). The scale is a six-point holistic scale, designed to measure the quality of each essay. A paper receiving a score of one does not control English grammar or syntax well enough to be comprehensible. A paper receiving a six would be virtually error free and comparable in quality to a paper written by an educated native speaker. The other ratings fall between these two extremes. For an exact copy of the guidelines used in the scoring of the TWE, and used in this study, see Appendix B. Sample essays with ratings are also included. For this study, the six-point scale was adapted to include plus or minus ratings, for example, 3+ or 3-. It was the opinion of the researcher that many students' papers might not improve enough from the first to second draft to merit an entire point increase. The scale was redefined to permit the demonstration of smaller changes in quality. For statistical purposes, a plus rating was 0.3 above the score and a minus rating 0.3 below the score (i.e., a 3+ = 3.3 and a 3- = 2.7).
The reliability of the TWE scoring method has been shown by ETS to be quite high, .80 to .85 in interrater reliability with scorers trained by ETS. It was not possible within the confines of this study to use scorers trained by ETS, nor did the topics used in this research necessarily correlate with any topics used by the TWE. However, all the raters used in the study were familiar with the procedure used by ETS.

For the TWE rating, a training session was held for all the raters to establish the common criteria necessary to rate the essays. Essays accumulated during the pilot study were used for this purpose. Each essay in the actual study was rated by two raters. If their scores differed by more than one point, a third rater was used. The final score given to each essay was the average of the two, or possibly three, raters' scores. A third rater was necessary in only ten of the one hundred sixteen essays scored. This would lead one to expect that the inter-rater reliability would be quite acceptable. The inter-rater reliability for the papers receiving oral comments (both first and second drafts) was 0.51. For the papers receiving written comments it was 0.63. The raters did not know whether the papers they were rating had receive oral or written comments, and all raters had combinations of both, since both existed in any one assignment by any of the teachers. Therefore, there is no immediately apparent reason for the difference in these reliability correlations. The holistic scale itself is consistent with current theories of writing research, and has been demonstrated therefore to have good construct validity (Greenburg, 1986). It measures the
skill it claims to measure, writing, so has good face validity. To some degree concurrent validity was determined by comparison to the other measures used in this study, particularly the error ratio score, though this was not central to the research design. There was no attempt to determine predictive validity within the confines of this study.

Length: Word Count

The second rating was a word count. Though not all the essays were counted by the same individual, there was no attempt to establish inter-rater reliability in this rating because of the nature of the task. All the individuals who participated in counting the words were told to count all words. If students combined two words into one improperly, the raters were instructed to count this as one word. Also, if students incorrectly broke one word into two, the raters counted this as two words. Hyphenated words were counted as one word.

A word count used to determine length has extremely good face/content validity. No formulas were involved. Every word was counted; therefore it tested what it was meant to test; length.

Errors: Error Ratio Score

The third rating was an error ratio score. In this rating, only the first two hundred fifty words of each essay were used. Two essays were not at least 250 words long. However, since the rating was a ratio, it was possible to compute the ratio using the actual number of words in the essay as the numerator instead of 250. Each
error in the first 250 words was underlined and assigned a score of 1 or 2. A score of 1 was used for an error that did not interfere with meaning. A score of 2 was used for an error that did interfere with meaning.

The error ratio score was based upon the Composition Correctness Score developed by Brodkey and Young (1981). It has been adapted for use in this study. In their study, Brodkey and Young counted the first two hundred fifty words of each essay and dealt only with the errors in this section. They underlined each error and gave each a score from one to three. An error receiving a one did not interfere with meaning. An error receiving a three completely obscured the intended meaning. An error receiving a two fell somewhere between the two extremes. A ratio was then computed of the error score to the number of words (250) and the resulting score was the composition correctness score. In this study, as in theirs, only the first two hundred fifty words of each essay were used. Each error was underlined and given a score. However, for the purposes of this study, as mentioned above, there were only two scores, a one or a two. As in Brodkey and Young, an error receiving a one did not interfere with meaning; for example, misspellings that were clearly comprehensible, most subject-verb agreements, or other errors for which the context made the correct meaning absolutely clear. An error receiving a two did interfere with meaning; for example misspellings that created a new word (miss for mass) or were incomprehensible, incomprehensible syntax, or other errors that made it impossible to be absolutely sure of the intended meaning, even
with the help of context. A ratio was then computed of the error score to the number of words (250), and a resultant error ratio score obtained.

Brodkey and Young emphasize not giving raters too explicit a list of what types of errors should receive a score of one, two or three (i.e. all spelling errors = a 1). Their rationale is that guidelines that are too strict do not permit for the natural variability of errors; for example, sometimes spelling errors do not interfere with comprehension, while other times they make comprehension totally impossible. They found that though teachers may vary on specific error scores, overall scores for each total essay have good reliability.

To assertain reliability in this study, sixteen student essays were scored by two raters. Both raters were very familiar with Brodkey and Young’s Composition Correctness Score. The inter-rater reliability for the mean scores for the sixteen papers was 0.97. Inter-rater reliability was not determined for each individual error in each paper. The individual scores for each error do show variance between the raters. Brodkey and Young also found this to be true. However, because of the high reliability of the overall scores, the ratio was determined to have acceptable reliability. The remaining essays were scored by the researcher. Examples of essays rated by both raters are included in Appendix C.

The instructions given to the raters were taken and adapted from Brodkey and Young (1981) and are as follows:

1. Count the first two hundred fifty words of each essay.
2. Underline each error contained in this section; from spelling and mechanics through verb tenses, morphology, vocabulary, and logical connectives between sentences. Include every error that a literate reader might note. Entire phrases of five or six words may be underlined where no single source of error is readily analyzable (p. 160).

3. Assign a score of one or two to each error. An error receiving a one does not interfere with the readability of the essay. An error receiving a two does interfere with the readability of the essay.

4. Using the following ratio, figure out the error ratio score for each paper:

\[
\frac{250 \text{ (words)}}{\text{(sum of errors)}} = \text{error ratio score}
\]

The lower the error ratio score, the higher the percentage of errors per word. So a paper with a high error ratio score would be a paper with few errors. For example, a paper with an initial error score of 5 (almost error-free) would have an error ratio score of 50, and an initial score of 10 would have a ratio score of 25. As the number of errors increased, the ratios decreased in smaller and smaller increments. So a paper with an initial error score of 30 would have an error ratio score of 8.33, and an initial score of 35 would have an error ratio score of 7.14, only differing by slightly more than a point, compared to the 25 point difference between an initial score of 5 and 10. Despite this exponential decrease in scores, the ratios were used over the initial error scores for two reasons. In the first place, the difference was not expected to affect the results
significantly. Secondly, by converting the error scores to an error ratio score, the subsequent results were reversed, which made them easier to compare to the other data collected in the study. In other words, if left in their initial state, the error scores would decrease as the number of errors decreased. This would cause these scores to be opposite in order from the rest of the data. By using the error ratio score instead, these scores followed the same order as the other data.

As stated above, the sixteen papers were scored by two raters to determine inter-rater reliability. Inter-rater reliability was determined to be 0.97. The raters differed by less than one point on all but four of the essays. This is highly significant for papers with few errors, but as demonstrated above, as the numbers of errors in the error count increased, the error ratio decreased in smaller and smaller increments. Therefore if the raters differed by one point on a paper receiving a score of 9 or less, it would be a greater difference than for a paper receiving a score of 10 or more.

**Content: Percentage Scale**

A fourth rating was conducted, dealing with the changes in content from the first to the second draft. This percentage scale was developed by the researcher to determine how much the content, not the quality had changed from the first to the second draft. If a paper was completely different, it received a score of 100% change. If there had been no changes in the content, though perhaps grammatical errors had been altered, the paper received a score of 0% change. Papers which had varying amounts of changes in content
received scores between these two extremes. Except for the instructions below, the raters received no training in using this scale; however neither rater expressed any difficulty or confusion using this percentage scale. The instructions given to the raters for the measure of change in content are as follows:

1. Read the first and second draft of each essay together.
2. Determine what amount of the content in the essay has changed from draft one to draft two. This determination is done primarily holistically, not by counting sentences or paragraphs. The emphasis is strictly on changes in content, not the quality or changes in the quality of the essay.
3. Assign a percentage to the amount of content change. These percentages should be rounded to the tenth: 10% change, 20% change, 30% change, etc., but not 15% change, 22% change, etc.

Again, sixteen first and second drafts were used to determine inter-rater reliability. Since only one of the sixteen differed, inter-rater reliability had been established. The rest of the papers were rated by one of the two raters.
CHAPTER IV

FINDINGS

ANALYTIC TECHNIQUES

All data in this study were analyzed for correlation and for standard means and distribution using t-tests for both independent and paired variables. The statistical package used was Plot-50, developed by Tektronics, Inc., for use with their 40/50 personal computer system. All correlations in this study were Pearson's correlations. The first three ratings used in the study (the Test of Written English (TWE) scale, word count, and error ratio score, henceforth ERS) were correlated with each other first for the entire subject pool, then broken down into subgroups and correlated. There were four subgroups; the ratings of all the first drafts of papers receiving written comments, hereafter W.I, the ratings of all the second drafts of papers receiving written comments, hereafter W.II, the ratings of all the first drafts of papers receiving oral comments, O.I, and finally, the ratings of all the second drafts of papers receiving oral comments, O.II. Appendix D contains a list of all abbreviations used in this study. Correlations were made for all twelve variables (four groups of three variables each). For the purposes of correlation, the score of the percentage of content change was not included as it applied only to the second draft of each paper.
Additionally, descriptive statistics (mean, variance, standard deviation, standard error of the mean, coefficient of variation, skewness, kurtosis, minimum and maximum value, range, and median) were determined for each rating for each of the four above-mentioned groups.

At this point t-tests for equal variance were conducted for each of the groups, comparing all the means of the ratings for W.I with W.II and O.I with O.II. Finally, both a t-test of equal variance and a paired t-test were done on the means of the differences of the TWE ratings of W.I and W.II with O.I and O.II. These tests were also conducted on the difference of the means of the word count for papers with oral and written comments, the difference of the means of the ERS for oral and written, and on the difference in the percentage of content shift between papers with oral and written comments.

DESCRIPTION OF FINDINGS

Hypothesis One

The first hypothesis was that ESL students' compositions which have received oral comments during the writing process will show greater improvement than those which have received written comments, as determined by the holistic rating scale used with the Test of Written English (TWE) part of the Test of English as a Foreign Language (TOEFL). The findings gave qualified support to this hypothesis. The results of the t-test comparing the means of the TWE ratings for W.I and W.II showed no statistical difference in the means at the $p<0.05$ significance level. The results of the t-test
comparing the means of O.I and O.II for the TWE rating, did show a
significant statistical difference in the means, with O.II superior
to O.I. Therefore, on the basis of these two tests, it appears
that while the second drafts of papers receiving written comments
show no overall statistical improvement in quality from draft one to
draft two, those receiving oral comments show significant
statistical improvement in quality. The difference of the mean of
W.I subtracted from the mean of W.II was 0.19. For the mean of O.I
subtracted from O.II, the difference was 0.51. While the means of
the papers receiving written comments increased less than two
tenths of a point on the TWE scale, the means of the papers
receiving oral comments increased fully half a point in quality, as
determined by the TWE scale. Table III shows the results of the
t-tests.

TABLE III

t-TESTS FOR THE DIFFERENCE BETWEEN FIRST
AND SECOND DRAFTS FOR BOTH ORAL
AND WRITTEN COMMENTS
FOR THE TWE SCALE

<table>
<thead>
<tr>
<th></th>
<th>WRITTEN COMMENTS</th>
<th>ORAL COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN - DRAFT I</td>
<td>3.378</td>
<td>3.436</td>
</tr>
<tr>
<td>MEAN - DRAFT II</td>
<td>3.570</td>
<td>3.950</td>
</tr>
<tr>
<td>POOLED VARIANCE</td>
<td>0.510</td>
<td>0.551</td>
</tr>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>t STATISTIC</td>
<td>1.025</td>
<td>2.637</td>
</tr>
<tr>
<td>PROBABILITY LEVEL (P&lt;0.05)</td>
<td>P(t&gt;1.025)=P(t&lt;-1.025) = 0.155</td>
<td>P(t&gt;2.637)=P(t&lt;-2.637) = 0.005</td>
</tr>
</tbody>
</table>

When a two sample t-test for equal variance was done on the two
differences of the means, it showed that the difference between the
two was statistically significant at the $P<0.05$ level. However it was decided that a paired t-test would be more appropriate since the data were indeed paired (i.e., the data from each student's paper receiving written comments were paired with the same data from the student's paper receiving oral comments). The results of the paired t-test did not support the first hypothesis. This final test showed that when the differences of the means were paired, there was no statistically significant difference in the means. This means that individual students did not show a significant increase in quality in their papers receiving oral comments when compared to their papers receiving written comments. This may be due to the variance in the data. A Box and Whisker plot, showing the range, mean and concentration of the scores is included in Appendix E. When the groups were compared as a whole, there was a statistically significant increase in the quality of papers receiving written comments. Although individual student's papers did not show a difference, there was enough variance in the data to show a difference for the whole group. Therefore, qualified verification of the first hypothesis has been achieved. Refer to Tables IV and V for the results of the t-test for equal variance and the paired t-test.

Hypothesis Two

Hypothesis two has two parts: first, compositions receiving oral comments will show more changes at a content level, and second they will increase in length more than compositions receiving written comments. The analysis of the data did not support either part of the hypothesis.
### TABLE IV

**MEANS OF DIFFERENCES OF TWE SCORES FOR PAPERS RECEIVING ORAL AND WRITTEN COMMENTS USING THE TWO SAMPLE t-TEST FOR EQUAL VARIANCE**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean difference O.II - O.I</td>
<td>0.192</td>
</tr>
<tr>
<td>Mean difference W.II - W.I</td>
<td>0.514</td>
</tr>
<tr>
<td>Pooled variance</td>
<td>0.497</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>56</td>
</tr>
<tr>
<td>t statistic</td>
<td>-1.739</td>
</tr>
</tbody>
</table>

**Probability levels (P < 0.05)**

- $P(t > 1.739) = P(t < -1.739) = 0.044$

### TABLE V

**MEANS OF DIFFERENCES OF TWE SCORES FOR PAPERS RECEIVING ORAL AND WRITTEN COMMENTS USING THE PAIRED t-TEST**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of the differences</td>
<td>0.322</td>
</tr>
<tr>
<td>Variance of the differences</td>
<td>1.320</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>28</td>
</tr>
<tr>
<td>t statistic</td>
<td>1.510</td>
</tr>
</tbody>
</table>

**Probability levels (P < 0.05)**

- $P(t > 1.510) = P(t < -1.510) = 0.071$
Changes in Content. The mean percentage of content change from the first to the second draft for papers receiving oral comments was 32.759%; for papers receiving written comments, it was 27.586%, a difference of 5.172%. Again, both two-sample t-tests and paired t-tests were computed on the data. Although a visual comparison of the means did indicate that the papers in the oral comments group had undergone greater changes in content than the papers in the written comments group, when the differences were analyzed, they were not statistically significant. This is due, in part, to the wide variance in the results. The pooled variance was 766.256, and the variance of the differences was 1875.862. This wide variance overshadows the 5 point difference in the means. The subject pool was too small and the amount of variance in the results too great to get any statistical confirmation of the hypothesis. There also did not appear to be any statistical difference between the two tests, showing that the overall amounts of change in content did not differ greatly from the amounts of each student's change in content for the two papers (see Table VI). Additionally, histograms showing the variance and skew of the data are included in Appendix F.

Changes in Length. The second part of the hypothesis stated that papers receiving oral comments would increase more in length across drafts than papers receiving written comments. Analysis of the data does not confirm this hypothesis. Both groups did increase in length significantly, but not relative to one another. The mean number of words per paper for W.I was approximately 460 words; for W.II, it was 579 words, an overall increase of about 119 words per
TABLE VI

PERCENTAGE OF CONTENT CHANGE:
TWO SAMPLE t-TEST AND
PAIRED t-TEST

TWO SAMPLE t-TEST (EQUAL VARIANCE)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN CONTENT CHANGE - ORAL</td>
<td>35.759</td>
</tr>
<tr>
<td>MEAN CONTENT CHANGE - WRITTEN</td>
<td>27.586</td>
</tr>
<tr>
<td>POOLED VARIANCE</td>
<td>766.256</td>
</tr>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>56.0</td>
</tr>
<tr>
<td>t STATISTIC</td>
<td>0.712</td>
</tr>
</tbody>
</table>

PROBABILITY LEVELS (P<0.05)

\[ P(t>0.712) = P(t<-0.712) = 0.240 \]

PAIRED t-TEST

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN OF THE DIFFERENCES</td>
<td>5.172</td>
</tr>
<tr>
<td>VARIANCE OF THE DIFFERENCES</td>
<td>1875.862</td>
</tr>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>28.0</td>
</tr>
<tr>
<td>t STATISTIC</td>
<td>0.643</td>
</tr>
</tbody>
</table>

PROBABILITY LEVELS (P<0.05)

\[ P(t>0.643) = P(t<-0.643) = 0.263 \]
paper. The mean number of words for O.I was approximately 522 words and for O.II, 640 words, a difference of approximately 118 words. Thus, even though the second drafts of papers receiving oral comments were longer overall than papers receiving written comments, the first drafts of these papers were also correspondingly longer. Therefore, regardless of the types of comments used, the overall length of papers increased, and increased about the same amount. The medians of the amount of difference for each group were also quite similar. For papers receiving oral comments, the median increase in length was 73 words. For papers receiving written comments, the median increase was 69 words. There was one unusually long second draft (1474 words, 200 words longer than any other papers) that had received written comments and one paper receiving oral comments that decreased in length significantly (from 638 words long to 337) which may account for the lack of statistical difference. The variance in the difference of the means of O.I and O.II was 34,687. For the difference of the means of W.I and W.II, it was 20,801. Even if there were a difference in the means greater than one point between the oral and written groups, such a wide variance would require the difference to be substantial in order to be statistically significant (See Table VII). Histograms indicating the skew of these data are in Appendix G.

Hypothesis Three

Hypothesis three stated that compositions receiving written comments will show a greater decrease in errors of grammar and usage than those receiving oral comments. To test this hypothesis, an
TABLE VII

CHANGE IN LENGTH: DIFFERENCES BETWEEN DRAFT I AND DRAFT II

<table>
<thead>
<tr>
<th>DESCRIPTIVE STATISTICS</th>
<th>W.II-W.I</th>
<th>W.II-W.I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN DIFFERENCE OF DRAFTS II - I</td>
<td>118.690</td>
<td>119.414</td>
</tr>
<tr>
<td>VARIANCE</td>
<td>34,686.722</td>
<td>20,800.608</td>
</tr>
<tr>
<td>STANDARD DEVIATION</td>
<td>186.244</td>
<td>144.224</td>
</tr>
<tr>
<td>STANDARD ERROR OF THE MEAN</td>
<td>34.585</td>
<td>26.782</td>
</tr>
<tr>
<td>MINIMUM VALUE</td>
<td>-301.000</td>
<td>-89.000</td>
</tr>
<tr>
<td>MAXIMUM VALUE</td>
<td>547.000</td>
<td>692.000</td>
</tr>
<tr>
<td>RANGE</td>
<td>848.000</td>
<td>781.000</td>
</tr>
<tr>
<td>MEDIAN</td>
<td>73.000</td>
<td>69.000</td>
</tr>
</tbody>
</table>
TABLE VIII

<table>
<thead>
<tr>
<th></th>
<th>WRITTEN COMMENTS</th>
<th>ORAL COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN - DRAFT I</td>
<td>9.616</td>
<td>8.984</td>
</tr>
<tr>
<td>MEAN - DRAFT II</td>
<td>15.299</td>
<td>16.241</td>
</tr>
<tr>
<td>POOLED VARIANCE</td>
<td>71.555</td>
<td>74.056</td>
</tr>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>t STATISTIC</td>
<td>2.556</td>
<td>3.211</td>
</tr>
<tr>
<td>PROBABILITY LEVELS (P&lt;0.05)</td>
<td>P(t&gt;2.556)=P(t&lt;-2.556) = 0.006</td>
<td>P(t&gt;3.211)=P(t&lt;-3.211) = 0.001</td>
</tr>
</tbody>
</table>

The error ratio scores of the papers in the oral group improved 1.573 points over the error ratio scores of the papers in the written comments group. However, the medians reverse this order. The median of the increase for the oral comments papers was 3.490; for the written comments papers, it was 4.320, almost one point higher. This indicates that a few papers in the oral comments group decreased in errors dramatically, while the majority did not decrease as much. In the written comments group, single papers did not decrease in errors so dramatically, but the majority of papers did decrease in errors slightly more than those in the oral group (see Table IX). Because of the nature of the ratio, as mentioned in Chapter III, if the paper had few errors, a small decrease in the number of errors would show a marked increase in the ERS. A paper which had an error score of 10 on the first draft would have an ERS of 25. If, on the second draft, this paper received an error score of 5, the corresponding ERS would be 50. However, a paper with an
error score of 25 would receive an ERS of 10. If this paper's error score decreased by 5 on the second draft as well (to an error score of 20), its corresponding ERS would be 12.50, a difference of 2.5 points compared to 25 points for the same amount of decrease in errors.

TABLE IX

CHANGE IN ERROR RATIO SCORE: DIFFERENCES BETWEEN DRAFT I AND DRAFT II

<table>
<thead>
<tr>
<th>DESCRIPTIVE STATISTICS</th>
<th>O.II-O.I</th>
<th>W.II-W.I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN DIFFERENCE OF DRAFTS II - I</td>
<td>7.257</td>
<td>5.683</td>
</tr>
<tr>
<td>VARIANCE</td>
<td>107.114</td>
<td>50.832</td>
</tr>
<tr>
<td>STANDARD DEVIATION</td>
<td>10.350</td>
<td>7.130</td>
</tr>
<tr>
<td>STANDARD ERROR OF THE MEAN</td>
<td>1.922</td>
<td>1.324</td>
</tr>
<tr>
<td>MINIMUM VALUE</td>
<td>-5.100</td>
<td>-5.210</td>
</tr>
<tr>
<td>MAXIMUM VALUE</td>
<td>36.110</td>
<td>26.040</td>
</tr>
<tr>
<td>RANGE</td>
<td>41.210</td>
<td>31.250</td>
</tr>
<tr>
<td>MEDIAN</td>
<td>3.490</td>
<td>4.320</td>
</tr>
</tbody>
</table>

This accounts for the wide variance in these data, since a small decrease in errors could cause a large increase in the ERS. The variance of the differences of the papers in the oral group and those in the written group was 134.446, compared to the difference of the means at 1.573. Because the variance was considerably greater than the difference of the means, no statistically significant confirmation or rejection of the hypothesis was gained.

OTHER FINDINGS

It was stated at the outset of this chapter that correlations were run on all the variables in the study. The ratings of the TWE,
wordcount, and ERS were correlated for the entire subject pool. The only correlation with any significant positive correlation was that of the TWE and the ERS, at 0.576. The correlation of the TWE and the wordcount was 0.143, and the correlation of the word count and the ERS was -0.003. This indicates that there was no correlation between the TWE scores and the length of the papers; length was not considered an automatic indication of quality by the raters. The lack of correlation between the length of the paper and the number of errors indicates that students who wrote longer papers did not necessarily write papers which were more error-free, or vice versa. Nonetheless, the positive correlation between the TWE rating and the ERS does suggest that the raters used in this study were basing their ratings, at least in part, on the number of errors in the essay. Table X contains the correlation matrix for the error ratio score (ERS) and the Test of Written English (TWE) score. Correlations were also computed after the data had been broken down into the subgroups, W.I, W.II, O.I, and O.II. When divided thus, the positive correlation between the TWE score and the ERS was still in evidence in all but one case. The one anomaly was the correlation of O.I, the first draft of papers receiving oral comments. The correlation for the TWE and the ERS was 0.394. For W.I, this same correlation was 0.581, for W.II, 0.684, and for O.II, it was 0.575. Because of the divisions across raters, all raters rated some papers in each group, so it does not follow that this is the result of idiosyncrasies in the raters. Nor does it follow that this is the result of a different writing assignment, as all assignments had papers
TABLE X

CORRELATION MATRIX FOR TWE AND ERS FOR W.I, W.II, O.I, AND O.II

<table>
<thead>
<tr>
<th></th>
<th>W.I</th>
<th>W.II</th>
<th>O.I</th>
<th>O.II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TWE</td>
<td>ERS</td>
<td>TWE</td>
<td>ERS</td>
</tr>
<tr>
<td>W.I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWE</td>
<td>1.000</td>
<td>0.581</td>
<td>0.535</td>
<td>0.523</td>
</tr>
<tr>
<td>ERS</td>
<td>0.581</td>
<td>1.000</td>
<td>0.694</td>
<td>0.775</td>
</tr>
<tr>
<td>W.II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O.I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O.II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlations are positively significant at the 0.55 level.
receiving both written and oral comments. Nonetheless, either one of these reasons could be valid for one section of the data, and this section might be large enough to influence the rest of the data. Appendix H contains scattergrams of these correlations. These scattergrams make the general trends of the correlations much clearer.

Additional correlations showed that the TWE ratings for W.I did positively correlate with the ratings given for W.II (0.535) and the TWE ratings for O.I correlated significantly with O.II (0.533). Additionally, the ERS for W.I and W.II had a large positive correlation. The ERS for O.I and O.II also showed a positive correlation (0.48), but it was not as great as that for W.I and W.II. This can be interpreted to mean that while papers receiving written comments showed a definite and consistent decrease in errors from draft one to draft two, such a decrease was not immediately apparent with the papers receiving oral comments. It is appropriate for the variables to be correlated across drafts (W.I with W.II and O.I with O.II). A wide variance could indicate either that different raters in the TWE scoring were rating on different criteria, or that the students exhibited a great amount of variance in their writing, sometimes improving, sometimes reducing the quality. Since it has already been shown that the papers decreased in errors regardless of commenting style, and increased in quality significantly for papers receiving oral comments (and did not decrease in quality for papers receiving written comments), this correlation of the variables across the drafts is logical.
The correlations of the number of words showed no positive correlations to the TWE or the ERS in any of the subgroups. For W.I, the correlation of the TWE and the wordcount was 0.249, for W.II, -0.062. The correlation of the TWE and the wordcount for 0.I was 0.191, and for 0.II, it was -0.034. In the correlation of the wordcount and ERS, the correlations are as follows. For W.I, it was -0.024, for W.II, -0.221. The correlation for the wordcount and ERS for 0.I was 0.186, and 0.II, 0.195. This does seem to support the idea that length had no bearing when correlated with what the raters judged as quality, though lack of errors was considered to be an indication of quality. Table XI shows the correlation matrix for the wordcount.

<table>
<thead>
<tr>
<th>W.I</th>
<th>W.II</th>
<th>0.I</th>
<th>0.II</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWE</td>
<td>ERS</td>
<td>TWE</td>
<td>ERS</td>
</tr>
<tr>
<td>WORDS</td>
<td>0.249</td>
<td>0.111</td>
<td>-0.062</td>
</tr>
</tbody>
</table>

(The correlation must be greater than 0.55 to show significance).
CHAPTER V

SUMMARY AND DISCUSSION

The intent of this study was to determine the effects of teachers' comments, whether written or oral, on ESL students' writing. The results gave qualified support to the first hypothesis, that the use of oral conferencing improves the overall quality of students' papers more than written comments. However, the second hypothesis, that papers receiving oral conferencing would change more in content and increase more in length than papers receiving written comments was not conclusively supported. Moreover, the third hypothesis, that papers receiving written comments would show a greater decrease in errors than those receiving oral comments was not conclusively supported. However, it was found that the TWE ratings used to determine the first hypothesis correlated positively with the error ratio score used in the third hypothesis. This shows roughly that papers which had fewer errors in grammar, spelling, or syntax received higher TWE ratings regardless of content. In light of this correlation, the first and third hypotheses contradict one another. If indeed the TWE and error scores are correlated, then it would be incorrect to assume that one could increase while the other did not. In fact the results seem to support this idea of their interconnectedness.
CONCLUSIONS

The Teacher Variable

Perhaps the largest confounding variable in this study was the influence of the different teachers. It was impossible to control for this variable without severely interfering with the natural course of instruction. However, all conferences were taped and copies kept of all the written comments that students received. It was beyond the scope of this study to analyze the tapes and written comments made by the teachers, though this area certainly warrants a closer look. The task of categorizing and identifying natural discourse was beyond the time constraints of this study. The researcher listened to the oral conferences and read the written comments and made general descriptive notes on the conferencing style of each of the three teachers involved. A partial transcript of a conference by each teacher, as well as copies of papers with written comments by each teacher are included in Appendix I.

Teacher 1. The first of the three teachers differed significantly from the other two teachers in that she employed the use of a grader. This grader corrected the errors in grammar and sentence structure for all of the students receiving written comments. She was not involved in the papers receiving oral conferences. The grader’s comments were generally direct corrections, writing the correct form over or directly on top of the student’s error. Sometimes she simply pointed out the error, using a code to identify the type of error (ie., "sp" for spelling error). Explanations of these codes were not included on the student’s paper.
Occasionally the grader made general comments or comments about organization or content, but this was rare.

In her written comments, this teacher also showed a great deal of variance between the two assignments. The first assignment was a comparison/contrast paper. On this assignment, there were few written comments; one or two interlinear comments, and at most three or four final comments. These final comments were most frequently praise of or comments on organization. The comments throughout the essay focused on organization, faulty sentence structure, or grammar. In the conferences, again, the primary focus was on organization. However, the comments were much more specific, and she encouraged a great deal of input from the students in the conference, asking them what they felt about certain aspects of the writing. Nonetheless, the teacher did correct students' writing in the conference, or at least asked students if they could locate grammatical errors. Correcting grammar, in specific sentences, took up the bulk of the conference, as it did in the written comments (primarily done by the grader). The teacher also emphasized praise in both the written comments and in the oral conferences. However, in the oral conference, the student received several times the number of comments than those received by students with written comments. The students who received written comments had simple corrections of their grammatical errors, the students who received oral conferences also had these corrections explained if they were unsure.

The second assignment was a persuasive paper, outlining an argument and providing a solution. Again, the grader corrected the
grammar on the papers receiving written comments. This time, however, the teacher herself wrote much longer comments at the end of each paper. Most of these comments focused on the nature of an argument paper and what changes were needed to make their paper an argument paper. It was apparent that there had been a great deal more confusion with this topic than with the first topic. The conferences also reflected this difference. The conferences still dealt with a great deal of grammatical issues, but also dealt with the topic of argumentation considerably.

This first teacher had had experience and training in teaching writing to first language writers. The influence of this greater and more varied experience with writing instruction was obvious when compared with the other teachers. She was more adaptable than the other two teachers, and altered her conferencing or commenting style to suit the task at hand. Additionally, almost all of her students’ papers improved from the first to the second draft, both in quality and number of errors, regardless of the type of comments, oral or written, she used. This was not the case for the second and third teachers. Several of the students’ papers in these groups actually decreased in quality from the first to the second draft, again, regardless of the type of comment. However, it was impossible to draw any real conclusions from this because of the difference in size between the three groups.

Teacher 2. The second teacher wrote a greater number of comments throughout the paper than the first teacher, focusing primarily on content and secondarily on organization. Comments on
content were frequently attempts to elicit more information. Most comments were interlinear, in the margins throughout the paper, and only rarely were there a lot of final comments. There were approximately 1 to 3 sentences per comment, compared with the first teacher's 1 sentence. Although she did praise students' work, praise was not as much in evidence as it had been in the first teacher's comments. She occasionally dealt with grammar in her written comments, though most frequently she dealt only with spelling errors. In this case, she made a mark indicating that the word was misspelled, and left the student to figure out the correct spelling. In her class, papers were revised two times before they received a final grade. Grammar was dealt with between the second and third draft. Since the intent of this study was to alter the teacher's normal teaching style as little as possible, it was felt that this pattern should not be altered.

The content of her oral conferences and written comments was not significantly different. In both her conferences and her written comments, she frequently asked for more details or specifics about aspects of the topic. This teacher rarely dealt with grammar at all in her conferences, though she did occasionally deal with issues of grammar if they seemed critical to the comprehensibility of the piece. In only one of the nine conferences did grammar play a role (see Appendix I). There were almost no differences in her remarks between the two assignments given, although the two assignments seemed to differ greatly in the type of writing necessary.
The second teacher had been teaching ESL exclusively for over six years. She had received no special training beyond basic ESL methodology in the teaching of writing. She was also teaching the ENNR group, who seemed to differ from the ESL group in several ways, most significantly style of writing (their writing was far more informal, more like spoken English, than the ESL students' writing). On the first paper, in her class of nine students (who participated in the study), only two improved as judged by the TWE count. Both of these received oral conferences. The rest of the papers decreased in quality regardless of the type of commenting used. All of the papers increased in length, and all but two decreased in the number of errors. Of the two that increased the number of errors, one was in the oral group, the other in the written group. The one in the oral group was also one of the two which had had higher TWE scores. On the second group of papers, even though there did not seem to be a great difference in the style of commenting, all of the students showed improvement on the TWE scale. Again all increased in length and all but two decreased in number of errors. The two that increased in number of errors were not written by the same students who had done so in the first paper. Again, one of the two received an oral conference, the other written comments. There was not any significant difference in the amount of content change regardless of type of comments made or assignment.

Teacher 3. The third teacher wrote a far greater number of comments than either of the other teachers had done. Most of these comments dealt with content. Usually one paragraph of comments was
devoted entirely to content, typically trying to elicit more details or explanations from the student. A second paragraph dealt with one or two grammar problems that occurred throughout the essay. Most essays were also commented on interlinearly, with most of these comments focusing on corrections of grammatical errors. The total amount of comments (not including interlinear grammatical corrections) was from one-half page, to an entire page of writing, more than doubling the amount of written comments students received as compared to the other two teachers. Sometimes her comments were quite personal, relating experiences she had had that were similar to the one the student was writing about. She also generally praised every student at some point.

In oral conferences, this teacher also dealt with organization, but her primary focus was on content and grammar. In terms of content, she frequently tried to elicit more specific details and explanations as in her written comments. However, she seldom dealt with organization in the written comments, though it was often central in the oral conferences. As in her written comments, she made personal comments that related to the student's topic. Also, in her written comments, grammar was dealt with primarily by correction. This was true to some degree in the conference as well, but there was also a greater amount of explanation than in the written comments. Again, there was not a significant difference in the content of the comments on the different essays, although the second essay elicited a greater quantity of comments, both in oral conferences and written comments. The types of essays written were not significantly
different; both elicited types of descriptions; therefore, a
difference in style of commenting was not expected.

In this class, only four of the fourteen initial students
completed the entire study. Therefore, drawing any conclusions about
the types of comments used and their effect on this group is
impossible. For the first assignment given in her class, regardless
of types of commenting used, all of the students' papers improved as
judged by the TWE scale; all students' papers also increased in
length and decreased in the number of errors. All but one of them
changed significantly in content, from 30 to 50% change (the one
other paper showed no significant change in content at all). In the
second assignment, all but one of the students' papers improved on
the TWE scale. This one paper had received written comments, in fact
the greatest amount of written comments of any of the papers (a full
page). This paper also decreased in the number of words and
increased in the number of errors. The subject of the paper was very
personal, and the teacher had had a similar experience which she
discussed in her comments. One may conjecture that this loss of
quality was due to the difficulty in revising a piece of writing that
was so personal. Revising requires a certain amount of impersonal
detachment (see Elbow, 1981). This paper changed in content only
10%, though the comments requested extensive changes. One other
paper in this second group also decreased in length. This paper had
received an oral conference (see Appendix D) which focused primarily
on organizational issues. The shorter length coupled with the
increase in TWE score indicates that the student followed the
teacher's advice on tightening up her essay and removing extraneous or repetitive sections. Additionally, in the second assignment, only two of the papers showed any changes in content. Both of these papers had received written comments and both changed by only 10%. There was not any corresponding decrease in the amount of comments requesting changes in content; however in the oral conferences, both students justified their desire not to change the content.

This teacher had had similar training to the second teacher, and had not received any special instruction in the teaching of writing, nor taught writing to native speakers at any time. She had been teaching ESL for approximately 4 years.

On a purely descriptive level, with no attempt to statistically analyze the data, it appears that the two greatest differences between the oral conference and written comments are in quantity and student feedback. While the teachers frequently commented on similar items in written comments and in conference, they spent considerably more time on them in conference. In most written comments, a problem or issue was typically dealt with with just one sentence, occasionally two. In the oral conference, this same issue would receive several times the number of comments. In one conference by the second teacher, approximately eighteen of the teacher's comments discussed one incorrect word choice (see Appendix I). Similar problems in papers receiving written comments received only a few words or a sentence. This sheer number of comments seems to increase the likelihood that students would have a better understanding of what it was their teacher wished them to change in their writing.
The second difference was that students had the opportunity for feedback in the conference, the chance to ask the teachers to clarify their meanings and to justify their own reasons for writing something the way they did. The benefits from this seem obvious. The students were able to clarify their intended meanings and also clarify, through questions, what the teachers were asking them to do. In fact, in most conferences, the students did just that. Most conferences were a dynamic conversation about the text between the teacher and student. To comply with students' needs for privacy, the teachers had the option of turning off the tape recorders while the students were talking. Therefore, a complete collection of student input into the conferences was not made.

In Chapter II, it was noted that one risk that teachers run is the appropriation of the student text, taking control of the writing away from the student (see Sommers, 1982, and Brannon and Knoblauch, 1982). In this study, all of the teachers did just that, to varying degrees. The first teacher appropriated the students' texts the most, both in the interlinear corrections of the student's errors, and even more so in the context of the conference. The second teacher did so in both contexts equally, though some of her students seemed to resist handing over control to their teacher. However, it was these students' papers which decreased in quality. The third teacher appropriated the students' texts in the conference a great deal; however she seemed to avoid doing this in the written comments. Again, lack of appropriation did not necessarily increase the quality of the paper. This indicates that there needs to be some review of
the question of what teachers want their students to be able to do in writing. Appropriation of the text by the teacher may indeed increase the quality of the paper involved. However, it may not enable the student to carry this improvement over to the next paper. Lack of teacher appropriation may work in some instances, or help students to improve their writing over the long run, but the results suggest that it is not effective in immediately improving the quality of student writing. If, as was pointed out in Chapter IV, the quality of the writing is correlated with the number of errors, the reasons why appropriation of the text improves quality are obvious.

Other Confounding Variables

Another obvious confounding variable was the limited size of the study, particularly in the light of the influence of the individual teachers on the results. Twenty-nine students participated in the entire study. Since the study was repeated, the twenty-nine students received both written and oral comments during the course of the study. When these groups are broken down according to teacher, however, the sample pool becomes dismally small. The first teacher had sixteen students, the second teacher nine, and the third teacher only four who completed the study. Therefore, it is exceedingly difficult to compare these three groups and find anything significant to say about them. Additionally, the wide variance in the data collected for this small subject pool made it impossible to conclusively support or disprove any of the hypotheses.

The other confounding variables that should be mentioned are the different topics used for the essays; some topics may have been
easier or more difficult to revise or improve than others. One rater found the lack of definition of the topic for each individual paper to be a problem for the TWE rating as well. Though the varied levels of control of English exhibited by the subject pool should not have been a problem, since students were only being compared to their own revisions, it could possibly be a factor because of the students' differing amounts of ability to complete the intended revisions. Presumably more advanced students have received more writing instruction and therefore may be better able to revise. The first teacher, who taught the advanced group, mentioned that she was impressed by their ability to write and revise. Of course, work done in the classroom or in conference prior to the beginning of the study could also have affected the results of the study.

Tests and Measures

Some discussion seems necessary regarding the tests and measures used in the study. Since the TWE rating scale correlated so highly with errors, it was impossible to judge the quality of the content of the papers without the influence of errors. Several raters, familiar with the more traditional uses of the TWE rating scale, also mentioned difficulty using the scale with papers that had been written at home, possibly with help, and without the constraints of time. One student's papers were disallowed from the research because of apparent plagiarism. However, raters found the addition of a plus or a minus helpful in determining scores. In the TWE rating scale developed by ETS, only whole numbers could be used. Because it was
the opinion of the researcher that it would be extremely difficult for papers to move up an entire point on a six-point scale when rewritten, raters were permitted to put a plus or minus after the score. A plus equalled 0.3 above the score (i.e., 4+ equaled 4.3) and a minus was 0.3 below the score (4- equaled 3.7). This alleviated the difficulty in showing smaller amounts of improvement from the first to the second draft.

Additionally, the second rater used in the errors per word count commented on a problem inherent in the two point scale. With this scale a simple spelling error could be ranked equally with a severe grammatical error that did not significantly interfere with the meaning. Therefore students with poor but comprehensible spelling would receive scores as low as students with more complex problems with English grammar. Perhaps if the scale had been kept at the initial three-point differentiation, this could have been avoided, though the instructions given by Brodkey and Young do not imply this. The three-point scale used by Brodkey and Young was not used in this study because of the difficulty the researcher and the second rater (in a different study) encountered in trying to identify to which of the three groups an error belonged. If the scale had been adapted to rank simple spelling errors which do not interfere with meaning below grammatical errors, it is possible that the correlation between the TWE rating and the error scale might have been even greater. In addition because the scores are not clear indicators of the number of errors as the ratios become smaller, it may not be a good indicator of the true number of errors. Perhaps leaving the data in the
original score form, prior to computing the ratio, would have eliminated this problem.

IMPLICATIONS

Implications for Teaching

The implications of this study suggest that oral conferencing can be used to increase the quality of students' writing more than written comments, but that both can be used successfully. Although the results of this study are not conclusive, they certainly support continued use of and research into the use of conferencing for ESL students during the writing process. There may be no statistically significant benefit from the use of oral conferences, but the benefits mentioned above, of increased numbers of comments per issue and the opportunity for feedback certainly support the use of oral conferencing in ESL, even if for no other reason than the opportunity to practice English in the conference setting.

It has also been shown that native speakers consider the number of errors as very critical when judging the overall quality of a student's writing. This would imply that dealing with errors at some point in the writing process is critical to improving the quality of the paper. However, encouraging students to increase the length of their papers, does not result in corresponding increases in quality.

In order to decrease the errors in a paper, and thus most likely improve the quality of the paper, the teacher frequently may feel the need to appropriate the students' text. However, this may or may not influence later improvement. Since it has been shown that the
teacher has more of an influence on the improvement in student writing than any type of commenting used, it is tantamount that teachers be adequately trained to bear this responsibility. One implication for teaching is that teachers should look critically at their own style of commenting, and review the effects of it on their students' work. The content of the message received from the teacher is more critical than the medium through which the message received. Comments focusing on organization and error appear to promote greater improvements in the quality of student writing than comments eliciting more details or specifics. In addition, teachers must be flexible, and be able to adjust their style of commenting to suit the task at hand.

**Implications for Further Research**

The most obvious area in need of further research is the comments themselves. Research needs to be done to identify what types of comments teachers make, particularly in oral conferences. If the comments teachers make in conferences are shown to be very similar to those made in written comments, this would support the results of this study. With that, the effect of these different comments on revision also needs to be reviewed. In addition, research into the pragmatics of the conference would be helpful in determining what is helpful to students and what is not. Affective studies into the types of students who do well with conferences and those who do not is also called for. Moreover, videotaping of conferences would also aid in an affective study of the effects of
conferences. Initially, descriptive research needs to be conducted to categorize these elements, and only after this has been done should experimental studies be conducted.

It was beyond the scope of this thesis to analyze the wealth of data gained from the comments. Analysis and categorization are critical to improve the quality of writing instruction that ESL students receive. If it can be determined what helps the student most, in what situations, the teaching of writing would benefit from this immensely. It was the expectation of the researcher that oral conferences would show less appropriation of the text than written comments. This did not prove to be the case. To ascertain differences between the two mediums of commenting, perhaps training the teachers to use a specific form of oral conferencing would be in order. As long as the oral comments differ primarily only in quantity, not quality, from written comments, no perceptable difference between the two would be predicted.

Additional research also needs to be done into the revising process of students, what they feel or do when confronted with teachers comments. The amount of research into the effects of comments, particularly written comments for first language writers has increased dramatically in the last few years. Similar research replicating these studies with ESL students is badly needed.

Moreover, the longterm effects of teachers' comments on students' writing needs to be reviewed. This study found that appropriation of student text by the teacher did improve the quality of the paper. However, it may not help the student to become a
better writer. If one of the goals of writing instruction is to teach students to become effective revisors, then the effects of this appropriation must be carefully studied. In addition, if it is shown that appropriation of student text does not improve the student’s ability to revise, then some sort of study training teachers how not to appropriate students’ texts is needed.

More research also needs to be done looking into the relationship of errors and quality. If the primary determiner of quality is the number of errors, then a study reviewing ways teachers’ comments help students reduce errors should be done. In fact, a study looking into the whole issue of errors, academic literacy, and whether or not these are acceptable indicators of quality is in order. Where correction of error fits into process-approach writing instruction should also be reviewed.

In addition, this study needs to be replicated on a larger scale to be able to more clearly ascertain the effects of written comments and oral conferencing. It might also be necessary to replicate the study and give all the students oral conferences or written comments on the same draft, rather than splitting the group. This could help reduce the variance in the data. Since the teachers resulted in being the most critical elements in this study, replication of this study would warrant a significant increase in the number of teachers involved.
BIBLIOGRAPHY


Boiarsky, Carolyn, "What the Authorities Tell Us About Teaching Writing: Results of a Survey of Authorities on Teaching Composition," April 1984, ERIC ED 243 145.


Moran, Charles, "Reading Student Writing," Nov. 1982, ERIC ED 222 918.


APPENDIX A

INFORMED CONSENT

I, ________________________________, hereby agree to serve as a subject in the research project entitled "The Effects of Written Comments and Oral Conferencing on the Writing of ESL Students" conducted by Ann L. Kirk.

I understand that the study involves the use of written material, written during the regular course of events for my English as a Second Language writing class. I will be told prior to writing which assignments will be used in this study. In addition, I understand that one of my teacher-student writing conferences (in conjunction with these same assignments) will be audiotaped and that this audiotape will be used in the research. I also understand that if I consent, the investigator, Ann L. Kirk, may observe my teacher-student writing conference.

I understand that the possible risks to me associated with this study are risks of psychological discomfort from the invasion of privacy from having my writing studied by someone other than the individual for whom it was written and from having one of my teacher-student writing conferences audiotaped and the tape listened to by someone other than the individuals involved. However, I have been assured that strict anonymity will be maintained, at no time will my name be published.

It has been explained to me that the purpose of this study is to identify the impact on ESL writing of different types of commenting. The purpose is also to discover to some extent, what types of teacher comments promote successful revision strategies in students.

Ann L. Kirk has offered to answer any questions I may have about the study and what is expected or required of me in the study. I have been assured that all information I give will be kept confidential and that the identity of all subjects will remain anonymous.

I understand that I am free to withdraw from participation in this study at any time without jeopardizing my course grade or my relationship with Portland State University.

I have read and understand the foregoing information.

DATE ___________ SIGNATURE ________________________________

If you experience problems that are the result of your participation in this study, please contact the secretary of the Human Subjects Research and Review Committee, Office of Grants and Contracts, 303 Cramer Hall, Portland State University, 464-3417.
APPENDIX B
TEST OF WRITTEN ENGLISH (TWE)
SCORING GUIDELINES

Readers will assign scores based on the following scoring guide. Though examinees are asked to write on a specific topic, parts of the topic may be treated by implication. Readers should focus on what the examinee does well.

Scores

6  Clearly demonstrates competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.
   A paper in this category
   —is well organized and well developed
   —effectively addresses the writing task
   —uses appropriate details to support a thesis or illustrate ideas
   —shows unity, coherence, and progression
   —displays consistent facility in the use of language
   —demonstrates syntactic variety and appropriate word choice

5  Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will have occasional errors.
   A paper in this category
   —is generally well organized and well developed, though it may have fewer details than does a 6 paper
   —may address some parts of the task more effectively than others
   —shows unity, coherence, and progression
   —demonstrates some syntactic variety and range of vocabulary
   —displays facility in language, though it may have more errors than does a 6 paper

4  Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.
   A paper in this category
   —is adequately organized
   —addresses the writing topic adequately but may slight parts of the task
   —uses some details to support a thesis or illustrate ideas
   —demonstrates adequate but undistinguished or inconsistent facility with syntax and usage
   —may contain some serious errors that occasionally obscure meaning

3  Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level, or both.
   A paper in this category may reveal one or more of the following weaknesses:
   —inadequate organization or development
   —failure to support or illustrate generalizations with appropriate or sufficient detail
   —an accumulation of errors in sentence structure and/or usage
   —a noticeably inappropriate choice of words or word forms

2  Suggests incompetence in writing.
   A paper in this category is seriously flawed by one or more of the following weaknesses:
   —failure to organize or develop
   —little or no detail, or irrelevant specifics
   —serious and frequent errors in usage or sentence structure
   —serious problems with focus

1  Demonstrates incompetence in writing.
   A paper in this category will contain serious and persistent writing errors, may be illogical or incoherent, or may reveal the writer's inability to comprehend the question. A paper that is severely underdeveloped also falls into this category.

Papers that reject the assignment or fail to address the question in any way must be given to the Table Leader. Papers that exhibit absolutely no response at all must be given to the Table Leader.

Copyright © 1946 by Educational Testing Service. All rights reserved.
RATER 1

APPENDIX C

ERROR RATIO SCORE

ARGUMENT:

Lottery in Indonesia

Lottery in our society (Indonesia) is very common. Some people agree that lottery should be legal, but some people disagree about that. In other words, there are pro and contra about lottery in Indonesia. Should lottery be legal in Indonesia?

First of all, I would like to give the definition of lottery. Lottery is a game of chance in which people buy numbered tickets on prizes. We can find lottery in Indonesia for a long long time ago, maybe 50 years ago. Lottery used to play for having fun and usually people played for small value, but now people play from small value to big value.

Now, I would like to talk about pro people; people who agree that lottery should be legal in Indonesia. I think they have some reasons why they agree that lottery should be
First, lottery can change the level of someone's life, especially lottery is good for poor people. In this case, poor people have a chance to become rich if they have an opportunity to be a winner. In fact, I read in Sinar Harapan newspaper (a famous newspaper in Indonesia) two years ago that a farmer in Sumatra (parts in Indonesia) was a lottery winner for Rp 200,000,000 (C $125,000). After that, he bought a big house, a good car, and then he expanded his farm in Sumatra. In this case, lottery can change the level of someone's life; it can change to be better.

Second, the government can also get benefits from lottery. When people buy a lottery, the government takes tax from it. Also, when someone win a lottery, the government will take 10% of that money as a tax. For example, someone to be a winner for
From Rp 200,000,000, so, he or she will get the money around Rp 160,000,000, (± $ 112,500). In this case, the government gets benefits from lottery.

Besides there are some pro people, there are also have some contra people about lottery. First, it seems to me that some contra people have an opinion that lottery is bad for all people, especially for poor people. Because lottery can make people want to waste their money just for lottery. For the poor people or people are supported by welfare, lottery can make them worse. Because they will spend their money from welfare unwisely to buy lottery. May be, they will spend all of their money to buy lottery when they still don't get / win a lottery.

Second, lottery teaches people to like gambling. Lottery makes people lazy to work, because they just hope their lucky in lottery.
They just hope their money or income from lottery, they don't want to work anymore. At least, they become lazy to work. In fact, people can win $125,000 just in one day if he or she is lucky. If he or she has to work, for how long he or she can get $125,000? Also, in lottery people just put a small value of their money and if they are lucky, they can get a big value. As a result, people tend to be lazy to work. In fact, one of my friends who likes lottery very much. She always plays lottery everyday and she thinks that she doesn't have to work anymore because she often wins in a lottery. And the amount which she got from win a lottery can she use it for her living expenses for 2 months.

After I know about the reasons of some pro people and some contra people about lottery. I think, basically there are a lot
of benefits by legalizing lottery in Indonesia.
So, it seems to me that lottery should be legal in Indonesia. Because if I compare advantages to disadvantages of lottery, I find that there are more advantages of lottery than disadvantages of lottery.

ERS = 9.62
Lottery in Indonesia

Lottery in our society (in Indonesia) is very common. Some people agree that lottery should be legal, but some people disagree about that. In other words, there are pro and con about lottery in Indonesia. Should lottery be legal in Indonesia?

First of all, I would like to give the definition of lottery. Lottery is a game of chance in which people buy numbered tickets on prizes. We can find lottery in Indonesia for a long time ago, may be 50 years ago. Lottery used to play for having fun and usually people played for small value, but now people play from small value to big value.

Now, I would like to talk about pro people; people who agree that lottery should be legal in Indonesia. I think they have some reasons why they agree that lottery should be legal.
First, lottery can change the level of someone's life, especially lottery is good for poor people. In this case, poor people have chance to become rich if they have an opportunity to be a winner. In fact, I read 1 Smar Harapan newspaper (a famous newspaper in Indonesia) two years ago that a farmer in Sumatra (one of 1 parts in Indonesia) was a lottery winner 1 for Rp 200,000,000, (≈ $125,000). After that, he bought a big house, a good car, and then he expanded his farm in Sumatra. In this case, lottery can change the level of someone's life; it can change to be the better.

Second, the government can also get 1 benefits from lottery. When people buy a lottery, the government takes tax from it. Also, when someone wins a lottery, the government will take 10% of that money as tax. For example, someone to be a winner for
From Rp 200,000,000.-. So, he or she will get the money around Rp 180,000,000.- (± $ 112,500).

In this case, the government gets benefits from lottery.

Besides there are some pro people, there are also have some contra people about lottery. First, it seems to me that some contra people have an opinion that lottery is bad for all people, especially for poor people. Because lottery can make people want to waste their money just for lottery. For the poor people or people are supported by welfare, Lottery can make them worse. Because they will spend their money from welfare unwisely to buy lottery. Maybe they will spend all of their money to buy lottery when they still don't get / win a lottery.

Second, Lottery teaches people to like gambling. Lottery makes people lazy to work, because they just hope their lucky in lottery.
They just hope their money or income from lottery, they don't want to work anymore. At least, they become lazy to work. In fact, people can win $125,000 just in one day if he or she is lucky. If he or she has to work, for how long he or she can get $125,000? Also, in lottery people just put a small value of their money and if they are lucky, they can get a big value. As a result, people tend to be lazy to work. In fact, one of my friends who likes lottery very much, she always plays lottery everyday and she thinks that she doesn't have to work anymore because she often wins in a lottery. And the amount which she got from win a lottery can she use it for her living expenses for 2 months.

After I know about the reasons of some pro people and some contro people about lottery. I think, basically, there are a lot
of benefits by legalizing lottery in Indonesia.

So, it seems to me that lottery should be legal in Indonesia. Because if I compare advantages to disadvantages of lottery, I find that there are more advantages of lottery than disadvantages of lottery.

\[
\frac{250}{26} \quad \text{ERS} = 9.62
\]
RATER 1

One child in one family

The government of China allow people to have one child in each family to solve overpopulation. Then the media began to publicize the benefits of having one child in each family. The media said 'Can be well taken care of. He/She will grow up healthy. Also, the parents will have more time to work or study. The parents do pay all attention to take care their only child. They want them become a specialist and they have only one chance to have pride of their child. They arrange everything for their child. What happens to the children? Most of them are less independent than those whose family have more children. These children don't have companion to
play with. They feel alone, the parents have to spend much time to play with them.

Daring playing, the parents often reluctantly yield precedence to their child. Sometimes, these kind children become more carefree and difficult to get along with.

The big benefit that media said about one child family has one child is to decrease population, and therefore, reduce cases of abortion. Let's see what happen in thirty years.

In China, according to customs, old people are take cared of by their children. Government does not have responsibility to take care of them. In the future, when these parents, who have one child, get old, who take care of them?
At that time a young couple will have to take care of your old parents. They won't be able to do so, unless they won't work. If they don't work, they will not have money to live. This situation allows government to take care of these elders. It will increase cares of nation.

Although one child in one family caused decrease population, it also causes the problems about family and nation.

ERS = 0.10
One child in one family

The government of China allow people to have one child in each family to solve overpopulation. Then, the media began to publicize the benefits of having one child in each family.

The media said one child can be well taken care of. He/she will grow up healthily. Also the parents will have more time to work or study.

The parents do pay all attention to take care their only child. They want them become a specialist and they have only one chance to have pride of their child. They arrange everything for their child. What happens to the children? Most of them are less independent than those whose family have more children.

These children don't have companion to
play with. They feel alone. The parents have to spend much time to play with them.
During playing the parents often modestly yield precedence to their child. Sometimes, these kind children become more careless and difficult to get along with.

The big benefit that India said about one child family has one child is to decrease population, and therefore, reduce cares of nation. Let's see what happen in thirty years.
In China, according to custom, old people are taken care of by their children. Government does not have responsibility to take care of them. In the future, when these parents, who have one child, get old, who will take care of them?
At that time, a young couple will have to take care of their old parents. They won’t be able to do so, unless they work.

If they don’t work, they will not have money to live. This situation allows government to take care of these elders. It will increase cases of nation.

Although one child in one family caused decrease population, it also causes the problems about family and nation.

\[
\frac{250}{41}
\]

ERS = 6.10
APPENDIX D

LIST OF ABBREVIATIONS

EFL . . . . . . . . English as a Foreign Language
ENNR. . . . . . . English for NonNative Residents
ERS . . . . . . . Error Ratio Score
ESL . . . . . . . English as a Second Language
MTELP . . . . . . Michigan Test of English Language Proficiency
O.I . . . . . . . . Paper Receiving Oral Comments, Draft I
O.II. . . . . . . Paper Receiving Oral Comments, Draft II
TOEFL . . . . . Test of English as a Foreign Language
TWE . . . . . . . Test of Written English
W.I . . . . . . . . Paper Receiving Written Comments, Draft I
W.II. . . . . . . Paper Receiving Written Comments, Draft II
APPENDIX E

TWE SCORE: BOX AND WHISKER PLOT

TWE SCORES: BOX-AND-WHISKER PLOT

VARIABLE: W.I

TWE SCORE
APPENDIX F

CHANGE IN CONTENT: HISTOGRAMS

HISTOGRAM

% OF CHANGE IN WRITTEN CONTENT: DRAFT I TO DRAFT II

0 10 20 30 40 50 60 70 80 90 100

0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0
HISTOGRAM

CHANGE IN NUMBER OF WORDS: DRAFT I TO DRAFT II

FREQUENCY
Test of Written English Correlations
SCATTER PLOTS

ERROR RATIO SCORE

THE SCORE: ALL DRAFTS

R5AII **
L14BII *
L6AII * L17AII *
SCATTER PLOTS
APPENDIX I

Transcripts of
Teacher/Student Conferences

Teacher 1

T. okay, okay I started writing down notes, what did I do with it? ah here it is, I didn’t write anything yet.
    okay this was a nice paper XXX you’ve got a good topic to compare.

S. um

A few things made me a little confused. um one is this when you say ‘from having mothers in two countries I have learned some differences between the two.’ Now, what’s un, unclear here is are you gonna, are you talking about the differences between Japanese mothers and American mothers or just between these two people?

I think um if I say American mothers and Japanese mothers its really hard because my mother a little bit different from typical Japanese mothers.

Uhhuh

Sometimes it is hard even to, hard to generalize.

Well that’s right and I agree.

So maybe...

I agree

between two people.

I think you just gotta make that clear.

okay

When you.....the only reason it is confusing say ‘from having mothers in two countries I have learned some differences between the two.” it is a little ambiguous between the two countries or between the two mothers but from the rest of your essay I could tell you really wanted to compare these two people as separate individuals not so much as cultural representatives. Right?

Sometimes little bit I can tell little bit but, you know it so big country here and so many mothers...

And to make a good paper I mean to be as specific as possible will really help

okay
and so, you know if you want to say that, if you want to expand this at all and say that some of these things might be cultural you can but otherwise just keep it, you know, these, these separate individuals.

Uhhuh, okay

Okay, now this, now it looks like you could've, this is one paragraph here. Be careful about having a paragraph with just one sentence in it.

Oh, okay.

I would keep this with this paragraph

Up here.

Yeah, because this is still kind of telling about your plan of development, what you are going to do with it, there are some similarities and, but also I like what you say here but you got to reword this, XXX, 'cause it's just a little, it doesn't read very smoothly' there are a lot of similarities between them and that could make me get along so well with mom when I first met her from the beginning'. You might want to use a because clause or something, 'because of their similarities I felt very

Oh

comfortable when I , maybe it was for this reason that it was so easy for me to adjust to life in the United States.' Something like that

okay

so it's a little bit smoother. But it's a nice sentence and I think it's important so let's put that with your introduction. Okay, then you start talking about your mother. You've got an interesting um plan of organization here and it works well. Now when you say Okasan works for Otasan you mean she works for your father in the home, right?

uh

Is that what you mean?

No, his store

She works for his store?

Huhhuh

Okay, I think you better make that clearer.
Okay

I wasn't quite sure what you meant by that, and you really don't have to put 'it means', let's just put 'father' in Japanese. Just like here, and also once you use the quotation marks I don't think you have to use them again, I mean, already you put up here 'Okasan' and 'mom' you've made the distinction between the two people, after that you can drop the...

Okay

Okay. Okay, so I'd make it clear that she works in the store. Okay, 'she also does all the housekeeping'. This is a little awkward 'she also cooks so good and cleans everywhere'. It just doesn't sound smooth. 'She likes', you don't want a 'to' with 'taking', 'she likes taking care of people so she has quite alot of friends and they usually'..........oh..'come to see her at night. We also usually'...Maybe you could find another word here. What else is similar to usually?

...Um...... um....

Well, I know, it's hard for you. How about often?

Often.

Often would probably work better in this situation anyway, it just means that many times you have people....Your writing is, your writing is really good, XXX, so I, I get very particular with you because I know you want to work out some of the rough spots. So, these are little things but they are just things that will improve your writing I think.

Okay.
Teacher 2

I. Okay, here, this word fell. Do you mean fall down, XXX, I don't think you do, do you?

S. I felt... out...

You mean...

Feeling

Oh... I don't know if that's what you do mean... feeling

I always wanted (mumble, mumble)

Don't you mean that you didn't succeed?

I felt, yeah, I'm sorry. I felt....

How do you..

... in the rocks.

Oh!... Oh! you mean you...

... fell...

Falling, like falling off a rock?

Yeah.

You don't mean, actually like, in an examination you pass or you.....

I failed.... it means I don't, didn't pass.

You didn't pass, uh, which is....

Is failed, em? Failed.

How do you pronounce that?

F.... failed?

Failed, yeah.

Failed.

You see, you've got three words that sort of sound the same.
(Thanks X, this is the homework)
Failed...failed...oh failed.
You've got feel, foal,... okay, fail. Okay?
Huhuh.
You've got three different meanings. What's the past of feel?
Felled.
Mmmmm...
Feel?
Yesterday I..
Felt
had..
Felt
F..
Felt
Yeah, and what's the participle?
Felt.
Huhuh, the same. Okay, now yesterday down stairs I..
Felled.
Uhhhh, do you need the 'ed' on it?
Fell, no.
Huhuh, and what about I have.....
Feel.
Mmmmm...
Fell?
Mmm..
(Laughter)
Here, here you go..
Fallen?
Huhuh
Alright......
Then, then if you are talking about...
...fallen...
an examination it's fail...
Failed?
Huhuh, and the participle again is..
Failed.
Huhuh, they all sound very similar. But I think that's what you mean, isn't it? That you didn't succeed.
Yeah.
Okay, ....and uh........so it seems that your thesis comes all the way in this part here.
Uh, too much?
I don't know. Often when you use a process essay you don't really need to explain each step. With other essays you can. But with the process sometimes all you need is this one. "There are only a few steps to stop smoking." I think that's all you need. Then you can use these sentences as
(mumble)
your topic sentences. Do you see what I mean?
Okay.
So you can use "at first you really have to want to stop smoking". That could be your topic sentence....
Okay, so those...
..for paragraph one and then this can be your topic for paragraph two. You don't actually need..
Okay, alright.
This is so long that it loses its impact. Do you see what I mean? You start thinking uhhhhhhhhhhhh... It takes the reader away from your point a little bit. Um..

Okay, how should I say that in the thesis?

I think this thesis is fine here. And in a sense I would keep that as your thesis and finish your paragraph there.

Do I have to say three things?

You don't have to.

Just have two steps, that's it? Oh.

You don't have to. A lot of other essays you'd put the three things in but with a process, because there are so many steps you don't need to. You just say these are the steps.

Okay.

And then you can divide it however..

Okay.

So your first section is techniques to help you stop wanting to smoke? Is that it?

Yeah. You have to want it first.

MmmmmMmmmm

Because if you don't want it to stop you don't, you can't. I mean you have to think about it.

Okay, so it seems like you've divided it into three parts. It's the sort of before...

Huhuh.

..sort of the during and after. Is that sort of how you've done it?

Huhuh.

Okay.
I. Um....I wonder if we should put this up a little bit higher so we don't have quite so much background first. Um, we might say something like...um......my......okay so one and a half year, this was the time. It was this one and a half year period that you you're talking about. Okay. So maybe right up at the beginning we should say .....um..." When I was in second grade in high school I decided to apply to the musical university. I had one and a half years of, years left. Right? After you decided? Is that right?

Yeah.

You decided then and then you had one and a half years and during that time you learned to be patient and hardworking. Or you realized that you were patient and hardworking. So maybe a little bit of explanation right at the beginning, okay..."I decided to attend. I had one and a half years to study and practice during that time." And then give more of the explanation. Okay, so right at the beginning I know what's it about. Otherwise I have to kind of struggle through the whole thing......she mentions this and this...why is she telling me all this?

Huhuh.

So maybe move the...up to the beginning.

Okay.

And, yeah, I wouldn’t, I wouldn’t say "not so long some time ago just one year ago". I would say "one year ago I tried".

Okay.

In the beginning right there.

"I tried." And take out the period. "I tried to attend." Um, now this is really, this came first and then this. Right?

Uh..

So you might want to say something like " when I was in second grade in high school I decided to apply to the Musical University. In that time I had one and a half years left for, at that time, I had one a half years left to the entrance examination. One year ago I had the examination." Okay, or maybe something like "two and a half years ago, when I was in second grade in high school, I decided to apply to the Musical University." So that we know and then um, "I had one and a half years left". Okay, " during ...that time ...." okay, then we are talking about during, during, during the one and a half years. Okay, during that time. Instead of noticed let's say realized. A
little bit stronger word there. "I realized that I became a patient and hardworking person." Okay, and then let’s go into the explanation okay, so that you’re...Now, now I see, okay, this is why she is telling me all this. She’s telling me because she is explaining how she was patient and hardworking. Okay? These are, these are some of your examples of your patience and hardworking. Okay? (teacher reading to themselves) Okay, it was in September. We might, it might help if you talk about some dates here. Maybe we should do that. In nineteen...

Huhuh.

...when was the date? When, when was the ...when, when did you decide to apply? That might make it a little bit clearer.

Um, in June...

In what, what year? (pause) (laughter) Three years ago? Two and a half years ago? 1985?

Eighty, eighty-five.

It might help, like in September here. I don’t know if this was September, 1985, or a year later or...

Um..

The time is a little bit confusing. So, let’s see. If you start out this thing "In 1985, when I was in second grade in high school I decided to apply". And then in 1980, in June, 1985 you decided to apply. When was the test?

Um, nineteen, oh, nineteen eighty-seven. Just last week, last year.

Okay, so maybe then you could say "In 1987 I had to apply". Then you can...and then kind of...."my last spring break".....maybe if you, then if you kind of talk about the months I can figure out what, what’s going on.....

Yeah.

....a little bit more. Okay, I’m just trying to think of some different ways to, so that I can, I’m, I’m a little bit confused. For example, here’s the September, and I think well when in September. Right away your six months later or a year later or...you know, it’s a little bit confusing. Okay, um...okay, at that time, now at what time? At the time that you decided to apply?

Yes, I decided.

Okay, let’s see if we can...Okay, so "In June, 1980...?...
1985

"1985 I decided to apply", with a 'l', (laughter) "to Musical University". Okay. "I had one and a half years (conference drowned out by small chorus singing 'Happy Birthday")

Um.."to prepare." Right? (more laughter) Okay. "One and a half years to prepare myself." Maybe, or, I'm just kind of writing my ideas down here. Okay. Okay, "then during this time ....blahblahblah..." Okay, " I realized I had become patient and hardworking." Okay. Now, um, you might want to say something, that my decision was unusual or my situation was unusual. Okay? Because I'm not sure what was unusual. My situation was unusual and then this explanation. Okay?

Oh.

And then maybe explain here, "In September 1985", okay, and so then I...I now that this happened in June, you had one and a half years that this was, now it's even less than one and a half years, okay, September 1985. Okay. "I did a lot", okay, work is non-temp, a lot of work at that. Now here, we can say during this time, during is followed by a noun. Okay, so this is okay, "During class in high school"...okay...

Huhuh.

"I did a lot of work during that one and a half year."

Huhuh.

Okay, "During class in high school I studied....(mumbled reading) You need to watch your 'l's and your 'r's.

Oh.

(laughter)
Graduating People from College and High School in Jakarta.

There are some advantages of graduating from college and
advantages for people who drop out of college, and some
some disadvantages of graduating from high school in Jakarta.

Most people in Jakarta will perhaps continue their studying in
to work. In this circumstance, if their parents are still able to support them. However, if their parents are not able to support them anymore, they will probably have to stop going to school. In this circumstance, most of them choose to start working.

In Jakarta, it's a good idea to continue studying in
college after graduating from high school. The reason is that people who graduate from college will have a higher salary. If they go to work, most companies will pay them with a higher salary, because they have a degree from college and they have a broad knowledge of the fields they were majoring in college.

However, if people who have just graduated from high school go to work, they will always have a low salary because...
They don't have any degrees and they don't have enough education to prepare for an office job. "Work" is a better word, I think.

People with college degrees can easily find a job in the office. In college, they study a lot of material that relates to their jobs. They have prepared themselves to do office jobs after graduating from college. Therefore, most companies trust their ability to do the office jobs. It seems like there is no doubt for people who have a high GPA to find a good job easily in Jakarta.

On the other hand, high school graduates will have much more difficulty getting a good job. They only study general studies in high school, and some of them don't even really study hard. Therefore, most companies almost don't want to hire them for office jobs. Only a few of them who work in the office.

Another advantage is that people who graduate from college have so that they will probably have better abilities to analyze.
some problems than those people who graduated from high school.

In college, people learn how to solve more difficult problems than those in high school, and it might make them have to think more critically, more carefully, and more accurately. But, in high school, they have easier problems than those in college, and they don't have to study as hard as in college. That's probably the reason why they have lower ability to analyze problems than college graduates.

In my opinion, if their parents are still able to support them to continue their studying in college, they had better go for it. Because, in college, they can learn more knowledge than in high school. Besides, nowadays, almost all of the companies in Jakarta prefer to hire college graduates than high school graduates.

Good. This is easier to organize, I think. Some rough spots - especially for introduction.
Best and Worst Movies

"Indians have made many English movies at Cinema. One main difference is that "Indiana Jones" is an adventure and a Western movie named "Red Scorpion Man" with many English movies. The movie "Indiana Jones" was made by the United States, and the movie "Red Scorpion Man" was made by a Western movie named "Indiana Jones."
In a plane, they reach India by some means. There very help poor people and at the end they come back. They very simple story and arranged. Johnny is very well on the other hand. The movie Red Sonja has all very complex story, they have never have same dimension. Parker's want to rule them but the fight them. Ken he can meet her boy gather fight with kindness. 3 people in this way. I can say that movie start very poor and more.

The photography of The Indiana Jones movie very good. We can see a lot of scenes in this movie. We can see the Great Wall of China in this movie. Indian cultural photography is very well, among movie. Cameramen try to make a good scenery. In this movie, Cameramen try to make a really nice. And if you compare these scenes with another movie Red Sonja. You can feel a lot of difference. Between them, because this movie is made in color a single scene one no good scenery. In what way?

Workah — I see no differences between these pictures. How? If you have some idea like me then when you read this movie you can feel like me. This seems to be an interesting idea for comparison, but would switch the order of supports left to right.

Also try to be more specific. You say "Indiana Jones" has good photography of. China and India, but be more specific about particular places and camera techniques. Check your spelling.
SCUBA DIVING

When I was in high school, I began scuba diving. I have very beautiful memories of my scuba diving. I went to the beautiful ocean at the vacation then I enjoyed scuba diving. Sometime I stayed on the boat 2 or 3 days. Usually I played scuba diving near my house, but sometime I went to more beautiful island by airplane.

When I arrived sea, I smelled sea and salt from the sea. I went to the beautiful diving area by boat, when I ride on the boat, I feel very comfortable, because nice window touched my body. I began diving, so I jumped off from the boat, then I went to underground. At that moment, change the world. I feel very nice, like I become fish. As soon as I went to underground, my body feel cold, because enter the water inside the wet suit, but after couple of minutes, inside the wet suit became comfortable. When I diving, I can listen my breath or air. When I arrived bottom of the ocean, I can listen many kind of sound, for example, engine and strange noise from outside. I watch many coral and fishes. There are many kind of shape and
colors for example, gold or green fishes. Sometime I found rare goods, for example, old diver watch or big key. Usually I played diving about 1 and half hours.

Scuba diving makes me forget all the bad things. I want to continue play scuba diving. I never forget these experience.

I like the way you take the reader from the boat, to the dive, to the bottom of the sea. I wish you would describe a little more about what you saw at the bottom. How big were the fish? Were they strange looking? Did you see other animals, like jellyfish or octopus? Was it dark or light? How deep did you dive? Did you ever feel afraid?

Remember that this is all in the past. Check to make sure all your verbs are in the past tense!