OCCASIONAL PAPER #9
1985

THE DIFFERENTIAL EFFECTS OF SOURCR OF CORRECTIVE FEEDBACK ON ESL WRITING PROFICIENCY

Zhang Shuqiang

DEPARTMENT OF ENGLISH AS A SECOND LANGUAGE

UNIVERSITY OF HAWAII AT MANOA

O C C A S I O N A L P A P E R S E R I E S

In recent years, a number of graduate students in the Department of English as a Second Language have selected the thesis option as part of their Master of Arts degree program. Their research has covered a wide range of areas in second language learning and teaching. Many of these studies have attracted interest from others in the field, and in order to make these theses more widely available, selected titles are now published in the Occasional Paper Series. This series, a supplement to the departmental publication Working Papers, may also include reports of research by members of the ESL faculty. Publication of the Occasional Paper Series is underwritten by a grant from the Ruth Crymes Scholarship Fund. A list of available titles and prices may be obtained from the department and is also included in each issue of Working Papers.

The reports published in the Occasional Paper Series have the status of "progress reports", and may be published elsewhere in revised form.

Occasional Paper #9 is an MA thesis by Zhang Shuqiang. His thesis committee members were Craig Chaudron (chair), Michael Long, and Jack Richards. This work should be cited as follows:

Zhang Shuqiang. 1985. The Differential Effects of Source of Corrective Feedback on ESL Writing Proficiency. Occasional Paper #9. Honolulu: Department of English as a Second Language, University of Hawaii at Manoa.

ABSTRACT

In teaching writing to second language learners of English, various corrective procedures are employed as feedback to assist the revision stage of the writing process. These procedures include the use of peer, teacher and/or self-feedback as stimuli for successful revision. Whether or not any one of these types of feedback is superior to the others has not yet been determined objectively. Related to this question is how ESL learners themselves feel about those corrective procedures from different sources.

The experiment and survey reported here are intended to illuminate the differential effects of teacher feedback, peer feedback and self-feedback upon the informational/rhetorical and grammatical/mechanical aspects of ESL writing proficiency at three levels, ranging from lower-intermediate to advanced. It is found that across the proficiency levels, manipulation of the feedback type variable produces no significant differences on the informational/rhetorical aspect of ESL writing proficiency. But in the grammatical/mechanical category, feedback has a

main effect. There is also evidence to suggest that teacher feedback might be the moat effective procedure in dealing with grammatical inaccuracy. Survey results reveal that the supposedly "palatable" peer feedback is not as well received by ESL learners as the traditional teacher feedback.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

In teaching writing to second language learners of English, various instructional procedures are employed to assist the learner during the revision stage of the writing process. These procedures emphasize the use of teacher, peer or self-feedback as means of stimulating successful revision. But it has not yet been satisfactorily determined whether or not any one of these commonly adopted procedures is superior to the others as positive intervention in the revision process. Many books and articles have been written on the topic. In contrast, empirical evidence obtained from reasonably controlled studies is surprisingly scant. This discrepancy has already caused much concern in the ESL profession. As a result, ESL writing has increasingly emphasized quantitative studies so that the efficacy of these instructional procedures can be objectively verified.

The project reported here was conducted to verify the effects of corrective feedback from various sources. The research compares differences in revision improvements based upon feedback from instructors, classmates or individual student writers themselves. In addition to source of corrective feedback, the effect of learner proficiency is also analyzed. Both factors are combined in a 3 x 3 factorial design. Dependent measures include

impressionistic evaluations of discourse quality, and objective evaluations *of* formal linguistic competence. The learners' reactions to corrective feedback from different sources were also measured by means of a questionnaire. Their preferences are then compared with the results of the experiment in order to find whether or not learner preferences corroborate the objectively verified effects of these types of corrective feedback.

In keeping with the qualitative approach, the review of the literature on the topic and the report of the experiment and survey are concentrated on empirical evidence. Speculative theorization is thus relegated to a secondary role not by oversight but in the belief that unfalsifiable theorization is not likely to convince anyone who has already taken sides in a controversy. The comparative efficacy of teacher feedback, peer feedback and self-feedback has been a long standing issue. *It* is high time that attention and energy were directed towards more tangible research results for an objective clarification of the issue.

CHAPTER II

DEFINITION OF TERMS

A Working Definition of "Feedback"

Since the primary concern of this study is to examine the effects of different kinds of feedback on ESL written compositions, it is necessary to define "feedback" in the first place.

The most general sense of the term may denote any form of reaction or response that is perceived to be subsequent to and contingent upon a previous performance. However, more specifically, it refers to a process in which those factors that produce a certain result are themselves affected by the result. Here the important element in this more specific notion of "feedback" is the partial reversion of the effects of given factors to their source so as to reinforce or modify it. Such effects are often identified in ensuing performance, which can only be attributed to and explained by a combination of the initial factors and the feedback about the results of those factors.

These two definitions of "feedback", one static and the other dynamic or process-oriented, have interesting parallels in the literature that concerns the teaching of

composition. Their most representative counterparts are probably the concepts of 'feedback[R] in the works of Moffett and Lamberg. Moffett (1968) loosely defines feedback as "any information a learner receives as a result of his trials" (P.188), whereas Lamberg (1980) maintains that feedback should be defined as "information of performance which affects subsequent performance by influencing students' attention to particular matters so that those matters undergo a change in the subsequent performance" (P.66). The key word, he insists, is "affects". The crucial considerations are whether or not a response does have an effect upon the source and, in a learning situation, whether or not the effect is a desirable one. Moffett defines it so broadly that anything following a performance counts as feedback, which may be true, however, while Lamberg is only interested in defining feedback which makes a difference. Here we have a problem. For example, a teacher's correction coming after written errors is sufficient for Moffett's definition, but the correction may fail to promote the learner's second language development, and therefore is insufficient for Lamberg's view of feedback.

For the purpose of the present study, a working definition has been attempted as a compromise between Moffett's and Lamberg's *concepts.* Feedback is construed as any information that, in reacting to certain preceding performance, has the potential of affecting ensuing

linguistic performance by calling attention to inadequacies in the learner's interlanguage system, thereby facilitating remedy or refinement. Simply put, reactions that might influence subsequent performance are all considered as "feedback". The definition has expanded the scope delimited by Lamberg but still maintains a clear focus on the corrective potential of feedback -- an important defining aspect that is missing in Moffett's definition. This understanding of corrective feedback corresponds to Chaudron's (1977) conception of correction in teacher-student interaction as "any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner's utterance " (P.31). Although he mentions only teacher $correction_I$ peer corrections can be readily incorporated into his descriptive model of discourse in the corrective treatment of learners' errors.

In the context of teaching $composition_I$ naturally, these effects are projected to be positive in nature. Therefore, when researchers investigate the effects of corrective feedback upon ESL composition skills, they are particularly interested in effects that would help the learner to ratify or overcome errors and, at the same time, develop awareness of norms of correctness in the target language. So the beneficial, corrective potential of feedback is the focal point of this research.

Feedback from Three Sources

Feedback can be conveniently categorized into three types (Lamberg 1980, Partridge 1981): teacher feedback, peer feedback and self-feedback. Teacher feedback is defined as any stimulus for improvement supplied by a linguistically competent person, normally the teacher in a formal language learning situation. The teacher's criteria for judgment are assumed to be beyond question. Peer feedback, as the term implies, is supplied by individuals, usually classmates, who are comparable to the student writer in terms of overall linguistic abilities. Their input functions as stimulus for authentic negotiation between the student writer and his or her peer readers, gradually resulting in an improved text which communicates to the audience what the writer failed to convey in the previous effort. The concept of self-feedback, however, needs a little explanation. Speech, whether spoken or written, is directed chiefly to other people. Moffett (1968) argues that even when one purports to write for oneself, one cannot escape from the ultimately communcative consequence inherent in any use of language. Even in one's unspoken thoughts, it is as if one were addressing oneself. Thus, once beyond the moment of writing, the writer becomes the "other" person, and starts to feedback to himself Or herself. It is this psychological reality of "otherness" that constitutes the

basis for what is referred to in this paper as "self-feedback", i.e. judgments, insights or intuitions which the student writer generates while assuming the role of a critic and reviewing the text from some psychological distance. It is generally believed that, if successful revision is to take place, it must result from one or more of the three types of corrective feedback.

CHAPTER III

REVIEW OF THE LITERATURE

From the Product Model to the Process Model

The practice of writing has traditionally been seen as a sequential activity in which the task of recording ideas is completed step by step according to a rigid rhetorical plan. Since the writing activity is viewed as little more than filling in a prepared outline, the preoccupation is with a composed product (in such aspects as style, discourse, syntax and mechanics), rather than the composing process (in terms of how ideas are generated, refined, integrated and conveyed). Since the mid 1960's, as a result of some highly commendable work done in the field of teaching first language (L1) English composition (Braddock, Lloyd-Jones and Schoer 1963, Emig 1967, 1971, Murray 1968, 1972, Elbow 1973, Diederich 1974, Britton 1975, Shaughnessy 1977, Perl 1978, Young 1978, Flower 1979, Flower and Hayes 1979, Clifford 1981, Tate and Corbett 1981, Hairston 1982), a fundamental transformation has taken place in the understanding of writing. Writing is now seen not as the mere recording of pre-conceived, pre-sorted and pre-digested ideas, but as a dynamic and inventive process in which ideas may be discovered, reformulated, rejected or reorganized at

any moment during the immediate interaction between the writer and the evolving text. In this sense, the act of writing is understood as a facilitator of thought.

This change in the general understanding of composing, acclaimed as a revolutionary "paradigm shift" (Hairston 1982), has had a profound impact on the teaching of English as a second/foreign language. Following the trend in L1 research and pedagogy, ESL researchers and methodologists like Zamel (1976, 1982, 1983), Raimes (19791, Taylor (1981), and Watson (1982) maintain or suggest that the general principles of the process model should apply to non-native speakers as well. Their beleif has been reiterated by many practicing ESL teachers.

The Controversy over the Role of Revision

The new concept of writing has since translated into a pedagogical pattern involving pre-writing, writing and re-writing, which 'places composition revision in a central position', because *writing is a discovery procedure which relies heavily on the power of revision to clarify and refine that discovery' (Taylor 1981:5-8). The same view is expressed by Zamel (1982) when she maintains that revision should become the "main component" of composition instruction that recognizes the importance of generating, formulating and refining one's ideas. With writing viewed as a process, the teaching of writing becomes a kind of

*intervention ... in the process to improve that process or the product of that process" (Emig 1967:128). A logical question stemming from this notion of teaching presents itself: of the three kinds of feedback provided respectively by the instructor, peers, and the ESL student writer, which is the most conducive to development in the learner's composing skills?

Before we proceed to look for an answer to this question, we need to obtain a more comprehensive view of the so-called "central position" (Taylor 1981) of revision in the process model. There has been a controversy over the so-called "power of revision". On the one hand, there are empirical studies supporting the importance of revision. Fellows (1936) showed that L1 students receiving teacher corrections with a chance to revise improved more in grammar and punctuation than those without a chance to rewrite their compositions. More recently, Buxton (1958) asked one group of college freshmen to rewrite essays in response to teacher commentary. Another group also received teacher commentary but did not do any revising. The pre- and post-test scores revealed that revising improved the subjects' composing skills demonstrably as compared to those of the non-revisers. McColly (1963) conducted a study to test the hypothesis that "more writing alone means better writing". It was found to be untrue. He concluded that the activity in and of itself is fruitless, unless correction, revision

and preferably discussion of revision accompany the writing activity. These studies make a strong case for the emphasis given to revision in the process model*

On the other hand, there have been contentions against the multiple-draft writing task. The most frequently cited argument is that quantity counts far more than quality. "Language is a self-correcting and self-expanding system and the more it is used, the greater the facility there is in the use of it" (Erazmus 1960:301). In an experiment by Arnold (1963), one group of tenth graders were instructed to recompose their essays with reference to teacher feedback, while a second group did not revise. After a year, no difference was found in the writing performance of the two groups, suggesting that revision itself might be an insignificant factor in the training of writing skills. Corroboration of Arnold's finding is provided by Hensen (1978), who had one group of college students do teacher-guided revision and another group make only sentence-level corrections without rewriting their themes. The mean gains of the two groups showed no significant difference. Similar studies have also been conducted in ESL teaching. h classic example is Brière's (1966) pilot study in which revision was felt, rather than proved, to be detrimental to the shaping of target language behavior, e.g. fluency, appropriateness, adequacy and correctness. Brière himself was aware that his pre-experimental design plus uncontrolled confounds greatly undercut the validity of any claim he could make.

Nevertheless, he was willing to put quantity before quality, citing Roberts (1958) and Erazmus (1960) as his rationalization. In a review of such studies, Gorman (1979:190) points out "(Brière's) conclusions are, in my opinion, unjustified", because "there appears to be no way of deciding from the evidence provided in what measure either of the two methods contributed to the final result". Celce-Murcia (1974) recounted her experience with a "speedwriting" procedure which required no revision, allowed minimal feedback, and stressed exclusively the amount of writing to be produced. The crucial question of whether speedwriting had enhanced her students' ability to communicate went unanswered, But she felt "intuitively" that "the answer to *the* above question is 'yes'" (P.69).

In spite of the controversy, revision with feedback has remained a major component in the practice of composition instruction. Revision has been incorporated in both product-oriented and process-oriented instruction, although given a distinctly different role in the former than in the latter. Those teachers who regularly employ revision techniques -- they constitute the overwhelming majority -- are interested in the question: Which type of feedback can best expedite positive intervention at the rewriting stage?

The Emergence of Peer Feedback in ESL Teaching

Intervention can come from any of the three sources: teachers, peers or the writers themselves. Traditionally, it has been the teacher who is expected to provide the final feedback after the learner has incorporated his or her self-provided insights into the draft. The role of peer input, if any, has been kept to the minimum so as to prevent incompetent L2 users from "messing up" the purportedly unambiguous instructional input. More recently, for the purpose of exploring the dynamics in the writing process and assessing the relative efficacies of various feedback procedures, researchers in Ll writing have become interested in experiments with non-traditional correction methods. Meantime, interest in non-teacher correction, peer correction in particular, has grown considerably in the teaching of ESL. Today, although empirical experimentation is still scant in ESL and the results available are contradictory, peer correction has already become a common and important component in many ESL writing programs.

Most ESL commentators on the peer feedback issue agree that the recent institution of peer feedback is based not upon empirical evidence, but upon a rationale of Ll and L2 equivalence. Ll research from the 1960's on has been used as justification for L2 pedagogy. For example, Arapoff (1968:300) compares native learners with non-native learners

in terms of how grammar is learned. *She* reasons that "just as native speakers learn their language via a discovery and transformation process ... so too foreign students can learn to write via the same process". Jacobs (1982) describes how L1 and *L2* students cope with the same writing assignments. Her observations suggest that the problem of how to meet the requirement of a particular *writing* task transcends language factors and is shared by native and non-native speakers. And Edelsky's (1982) study *of* the L1 and L2 writings of bilingual children show that "general process universals" operate regardless of the language. It is this belief that prompts a transplant of L1 research results to ESL teaching and accordingly directs the ESL teachers[1] search for an efficient feedback in the direction of the hitherto unrecognized ESL peer audience.

Empirical Studies in L1 Research

The emergence of peer feedback as a technique for improving ESL compositions has its origin in the theorization and experimentation in L1 research. So far, speculations vary and research findings are inconclusive. The major findings of those L1 studies will be discussed in sections with respect to the effectiveness of individual types. It will be seen that there is evidence in support of each type as a contributor to writing improvement.

Research Results in Support of Teacher Feedback

Linn (1976) used the traditional product-centered teacher-dominated method and the innovative "free writing" method with two separate groups in a contrastive study. She was doubtful of the traditional approach and philosophically and psychologically more at ease with the new procedure. The "free writing" method initially freed the subjects from their writing anxiety and produced better essays, but in the long run proved to be not so effective as the traditional method. The results from post-test essays in favor of the traditional method had utilized teacher feedback as opposed to the peer evaluation in the new approach. It should be noted that there were several variables involved in the design, so it is not evident whether the gains could be legitimately attributed to the feedback factor.

Ziv (1981) had her subjects rewrite essays in response to both technical and rhetorical revision cues ranging from explicit directives to implicit suggestions, all of which were supplied by instructors. Her results show that teacher intervention does affect writing improvement in a multiple-draft assignment and has the potential to be a central activity of composition instruction concerned with stimulating and guiding revision. Her results also

challenge Knoblauch and Brannon's statement that teacher commentary on student essays is "an exercise in futility" (Knoblauch and Brannon 1981:1).

Haswell (1983) experimented with a marginal remark technique that had been devised to mobilize the learner's problem-solving potential. He also obtained a highly significant result to suggest that teacher feedback is by no means "an exercise in futility". Neither Ziv nor Haswell had a control group using an alternative feedback procedure. Therefore their findings can not substantiate any claim about the superiority of teacher feedback. But their results are impressive enough to warrant hypotheses about the positive effect of teacher feedback.

Apart from the studies cited above, there have been quite a number of studies that have reached *the* conclusion that teacher feedback, if it is *not* more facilitative, is neither more detrimental than the other two types of feedback. Such findings at least lend some weight to the argument that the traditional role of the instructor is not as futile or counter-productive as it *is* said to be (Erazmus 1960, George 1972, King 1979, Knoblauch and Brannon 1981, Marzano and Arthur 1977, Roberts 1958). Pierson (1967) compared three classes of ninth graders who received teacher feedback and another three classes who received peer feedback. He had hypothesized that the peer feedback classes would do better because of the importance they

placed on peer opinions. But scores on an essay test showed no statistically significant difference in improvement. However, this result was not included when the study was formally reported (1972) because the author had found the inter-rater reliability on the scores of the essay test too low to justify the claim.

A similar study was done by Farrell (1977), which investigated the comparative effectiveness of teacher feedback vs. peer evaluation vs. group tutoring by upper level students. While Farrell had posited that high school juniors tutored by seniors would improve the most on both an objective writing test and an essay test, he found no significant difference among the three groups. All improved more than students who had to utilize only their self-generated feedback.

Beach (1979) looked at the effects of between-draft teacher evaluation versus self-evaluation with or without a checklist. Differences in the effects were determined both in terms of the extent to which a rough draft was altered and the extent to which the final polished version differed from the first draft in quality. The quality scale covered focus (theme), sequence (organization), support (elaboration), sentence construction (syntax) and flavor (uniqueness, originality, vividness etc.). Again no significant differences emerged from all the quality scores

except in the category of support, where teacher guidance proved to be significantly beneficial, That means that, although teacher-guided students revised measurably more than the other groups, their essays were hardly any better than those of the other groups. The somehow blurred picture might be blamed on the rating scale. The validity of those impressionistic instruments had not been sufficiently established.

Pfeiffer (1981) arrived at a comparable conclusion in a study with college undergraduates. Peer correction and teacher correction did not cause any difference in writing performance. More interestingly, neither did they produce any difference in a measurement of writing anxiety. Pfeiffer's experiment is especially important because it questions the presumed "palatable feedback" offered by peers (Ellman 1975), thereby posing a serious challenge to the alleged affective advantage of the "collaborative writing pedagogy" (Clifford 1981) or other student-centered, process-oriented pedagogical models.

Research Results in Support of Peer Feedback

Interest in peer feedback as an alternative corrective approach has originated by and large from a long-standing sense of frustration with orthodox teacher correction. Such frustration is reinforced from time to time by studies like the one performed by Marzano and Arthur (1977). They had

three groups of learners receive respectively abbreviations indicative of error types, actual corrections and substantive comments designed to foster problem-solving processes. All the three types of guidance were supplied by instructors. No significant differences in effectiveness could be discerned. What is more discouraging is that all the types had equally small or insignificant influence on student writings, indicating that teacher intervention simply does not work.

Marzano and Arthur's morose conclusion was shared by King (1979). Dealing with grammatical accuracy alone, she discriminated three kinds of teacher comments: making explicit correction, naming error types, and offering relevant syntactic rules. She found that students rarerly understand what the teacher writes. Even if they understand, they do not or cannot implement the comments. Summing up such findings, Knoblauch and Brannon (1981:1) conclude that "positive results of teacher intervention through written commentary simply have not yet been found.' In other words, the time and energy invested are largely wasted.

Experiments such as those cited above do not make contrastive studies. For more pertinent information, better controlled experiments are required. For instance, Maize (1954) designed a project to test the hypothesis that more

effective results could be obtained if the teacher deliberately refrained from offering corrective feedback while at the same time he or she encouraged voluminous efforts by the class. One hundred and forty-nine college freshmen of low writing ability were randomly assigned to two groups. The control group wrote essays to be corrected by the teacher. The experimental group received only peer editing and peer evaluation, Then the two groups took a post-test in English usage and the subjects each wrote one controlled and one free composition for grading. On nearly all measures, the experimental group showed evidence of greater progress than the control. The researcher, however, found no difference between the two groups in their attitude towards writing and their writing instructors. (Also see Pfeiffer 1981 for *the* lack of identifiable affective advantage in peer feedback.) A very serious confound in the study was the unequal amount of writing practice for the two groups. The experimental group wrote forty essays as opposed to fourteen essays by the control group, which conceivably had biased the results in favor of peer feedback.

An experiment conducted by Putz (1970) centered around a comparison between "non-directive, student-centered learning" and "text-oriented, teacher-dominated learning" at the college freshman level. Comparison of pre-test and post-test scores revealed that neither group improved significantly. But the experimental (non-directive) group

did no worse, although it had not received any formal instruction or teacher commentary. Two of the confounds in the design, as Partridge has noted (1981), were the absence or presence of classroom instruction and the absence or presence of textbooks. The experimental group differed from the control not only because peer feedback was employed instead of teacher feedback, but also because they were not provided instruction or textbooks. Considering the disadvantages, peer feedback appears all the more superior to teacher feedback.

Ford (1973) produced evidence in favor of peer feedback. In a study with comparable college freshmen, two classes wrote six essays each. All the written assignments of one class were edited by instructors, and all the assignments of the other group were edited and evaluated by peers. The peer revision group performed much better than the control on an essay test as well as a post-test in grammar.

Sager (1973) explored the possibility of improving the quality of composition through the use of a rating scale in peer- and self-provided evaluation and correction. The control group relied exclusively on teacher feedback. All the subjects were sixth graders. Two classes used the rating scale to improve their own and each other's writings respectively. Their post-test writing samples were clearly

better than those produced by the control group, However, as Partridge (1981) points out, it is also possible that it was not the source of correction but the mode of input (rating scale vs. no rating scale) that made the difference.

Lagana (1974) worked with two tenth grade classes, one with teacher feedback, the other receiving peer editing and having conferences with the instructor. She found some differential effects with respect to content and form. The peer feedback group improved more in what can be called "higher order" concerns such as critical thinking, appropriateness and organization, whereas the teacher feedback group improved more in "lower order" concerns like spelling, punctuation and grammar. The findings should be taken with precaution because the design included more factors than just feedback type. The experimental group had actually received individualized teacher input during the conferences.

A quasi-experimental study by Karengianes, Myra and Pascarella (1980) investigated the influence of a peer editing treatment on *the* essay-writing proficiency of low-achieving tenth-grade students (writing at the seventh-grade level), Their post-test writing samples were rated significantly higher than *the* essays written by comparable low-achieving tenth-graders who had received teacher corrections. Both groups used a teacher-prepared checklist for self-evaluation in the course of rewriting.

Clifford (1981) developed a "collaborative composing method" stressing shared authority in the composition classroom and small group response as feedback at the revision stage. Ninety-two college freshmen were randomly assigned to an experimental group which was exposed to peer feedback, and a control group which received the traditional teacher commentary as final judgment. After a whole semester, the subjects' pee-test scores and post-test scores were tested (ANCOVA). The experimental group had significantly greater gains on the holistically scored post-test essays, but no difference was found in their performances in the grammatical or mechanical aspect of writing, inspite of the fact that the control group had received explicit classroom instruction on grammatical and mechanical points.

Besides quantitative studies, case studies were sometimes conducted to determine the effectiveness of peer feedback. Calkins (1978) described how eight- to nine-year-old pupils improved upon their drafts through group discussion. Here Calkins is cited not just to provide information about peer feedback, but also to alert the reader to the limitations of a typical case study, which is still perhaps the most commonly adopted procedure in describing the writing process. Reviewing Emig's prototypical case study (1971), Voss (1983) notes "the general prestige of science in our society has not been

earned by the highly inferential procedures of case study research . . . we should be more cautious in our extrapolations and interpretations of its results" (P.279). However insightful a case study generalization may be, its dependability is necessarily tenuous.

Research Results in Support of Self-Feedback

What we know about self-feedback has mainly been gathered from contrastive studies involving a control group without any input from either instructors or peers. Such information has already been included in the discussion above and will not be repeated here. The reader is referred particularly to Beach (1979), Farrell (1977), and Sager (1973).

One study gives strong support *to* self-feedback. In Wolter's (1975) experiment, he instructed one group *of* subjects to measure their own writings and another group to recompose under the guidance of teacher comments. The two groups did equally well at the end of the experiment, indicating that self-feedback is at least as effective as teacher feedback. The pedagogical implication is that learners can progress without or in spite of teacher intervention.

An "Oddball" Finding

Of the eighteen studies surveyed above, seven can be interpreted as supportive of teacher feedback, ten in favor of peer feedback and one in favor of self-feedback. But it should also be mentioned that most of the studies have confounds. And some of the results are equivocal. Apart from the eighteen studies, there is one study that does not fit under any of the three preceding headings. It is a one-of-a-kind study due to its peculiar discovery. It was posited in Sutton and Eliot's (1964) study that learners who evaluated others' themes would demonstrate more improvement than those who were passive recipients of correction. It was also hypothesized that peer feedback would cause more improvement than teacher feedback. Post-test scores showed that all the subjects, instead of gaining from feedback, declined in writing proficiency. If this strange finding were to be generalized, any form of feedback, including self-feedback, which the control group used, would debilitate writing competence. It is hardly conceivable that a learner can make progress without any guidance or even response. It is suspected that the administration of the experiment perhaps had failed somewhere to meet the rigor required by such quantitative studies.

Empirical Studies in L2 Research

Three Studies

To the knowledge of the writer, the only three experimental studies on this topic in ESL were conducted by Partridge (1981), Chaudron (1984), and Zhang and Halpern (1904). Because of the limited number of studies done and also their direct influence upon the study reported in this paper, these studies deserve more detailed discussion than those in Ll research.

Partridge (1981) had a group of twelve intermediate level ESL learners write six Compositions over a period of approximately six weeks. Compositions 1, 3, and 5 were corrected by instructors. The remaining three were evaluated and corrected through group discussion and peer commentary. After correction, the compositions were rewritten and then graded by two panels of raters according to an analytic scoring scale based on the model developed by Cooper (1977) for Ll learners. The scale specified criteria for an impressionistic measurement of three aspects of ESL writing: grammar, vocabulary and style. The last category, style, in Partridge's study did not include subcategories like effectiveness, sincerity etc., which were originally in Cooper's scale. In the course of the experiment, the subjects were also asked to provide information about how

they felt about peer feedback. Although the students' reaction to peer feedback was generally favorable, results of matched t-tests suggest that teacher intervention is more effective than peer intervention in improving the overall quality of ESL compositions.

Partridge's study has serious defects in design, statistics and measurement. Firstly, as the same group of learners were alternately exposed to teacher and peer feedback on a prolonged week-by-week basis, it is virtually impossible to determine, through the statistical procedures she employed, whether any observed progress could be traced back to particular sources of intervention. That is to say, the effects of the feedback from two identifiable sources had been pooled together through an on-going accumulative process. Secondly, isolating effects in a time-series design is not impossible, but the t-test is obviously a questionable method for the design. Finally, the reliability of measurement poses another problem. The six judges in two panels scored all the six assignments of the two groups. In all, thirty-six Pearson product-moment correlation coefficients are reported (3 pairs of raters x 2 panels x 6 assignments), and they range widely from - 0.89 to + 0.94. Of the thirty-six coefficients, only three are significant ($p<0.05$). The three significant coefficients have turned out to be at the opposite extremes of the range, - 0.89 at the lowest end and + 0.94 and + 0.93 at the top. No consensus is evident among the six raters. In very

straightforward language, those scores reflect six idiosyncratic rating scales, rather than a consistent evaluation. There is simply not a credible numerical basis for statistical inferences. In short, *the* value of Partridge's study lies more in the fact that it is one of the earliest *studies* on the topic than in any specific results it has yielded.

Chaudron (1984) performed another experimental study with one group of fourteen high-intermediate ESL learners and one group of nine advanced learners. Of the four out-of-class essays assigned, the middle *two* were experimental exercises, in which half of a class received peer evaluation as a basis for revision, while the other half received only teacher feedback. The *two* halves were reversed on the next assignment. The evaluation focused on both grammatical/mechanical errors and content/rhetorical weakness. The drafts and revisions were graded by independent judges using the ESL Composition Profile developed by Jacobs, Zingraf, Wormuth, Hartfiel and Hughey (1981). The Pearson product-moment correlation between the judges is significant ($p<0.001$). T-test comparisons show no overall difference between the improvement due to teacher feedback and its counterpart due to peer feedback. Student response to peer evaluation appears to be appreciative but cautious.

A more interesting but not fully explored part of the study is the discussion on the variation in the relative benefit students might derive from various sources of feedback. Chaudron noticed that the advanced group made progress with either treatment, whereas the high-intermediate subjects hardly improved, suggesting that proficiency level might be interacting with the feedback variable. His display of subcategory scores: content, organization, vocabulary and grammar points to the possibility of different feedback types exerting different influences upon different aspects of ESL writing. Unfortunately, the rather limited data base prevented more rigorous analysis.

Zhang and Halpern (1984) followed up the Chaudron study with a 2 x 3 factorial design study involving two independent variables: level of proficiency and source of feedback. The former had two levels: intermediate and advanced, the latter had three levels, namely, teacher feedback, peer feedback and self-provided feedback. The dependent variables examined included two major categories: content/discourse adequacy and grammar/mechanics accuracy. The content/discourse aspects were measured according to the criteria specified in Jacobs et al.'s ESL Composition Profile (1981), and the grammar/mechanics aspects were scored with objective frequency-based instruments borrowed from Homburg (1984). The Pearson product-moment correlation coefficient on the content/discourse measure is 0.80,

indicating acceptable inter-raker reliability. The percentages of agreement are 87.3% for the mechanical errors/T-unit scores, and 85.9% for the grammatical errors/T-unit scores. The results of an analysis *of* variance show that on either level of ESL proficiency, treatment has no effect on the content/discourse aspect of writing, but it does affect the grammar/mechanics aspect of ESL writing. Teacher feedback is generally superior in reducing mechanical or grammatical errors. Advanced learners working on their own made more or less the same progress as teacher- or peer-guided students. Lower-level ESL learners were relatively incapable of worth-while self-feedback. Only teacher correction measurably enhanced their linguistic accuracy. There is also tentative evidence to suggest that peer intervention could be valuable for enhancing the level of grammaticality. *An* interaction effect is found in the grammar dimension, indicating that the effects of feedback treatments are not independent of the effects of proficiency. This conclusion agrees with some *of* the findings in L1 research (Beach 1979, Lagana 1974).

One weakness of the study *is* that, because of the difficulty of breaking down the holistic measure of content/discourse adequacy, the question of how the feedback variable Influences improvements in content, organization and vocabulary separately had to be left unexplored,

A Natural Progression

An interesting observation emerges as these three studies are examined from a chronological perspective. During the four years from Partridge's study (1981) to Zhang and Halpern's study (1984), empirical research on this specific topic has demonstrated a very healthy trend. The sample size has been increasing from 12 subjects in Partridge's study to 23 in Chaudron's study to 62 in Zhang and Halpern's study. The design is improving too, involving more and more variables. The Partridge study looked at only one experimental variable and one dependent variable. Chaudron investigated chiefly the relationship between one independent variable and one dependent variable, but also included discussion on the possible effects of another factor, i.e. proficiency level, and the possible variation among subcategory scores, content, organization, vocabulary and grammar. Zhang and Halpern took the research a step further by adopting a factorial design which isolated the effects of two independent variables, namely feedback type and proficiency level, and their interaction effect upon four dependent variables. Meantime, more attention was paid to measurement as well. Partridge's conclusions are based upon raw scores that exhibit no trace of reliability. The inter-judge reliability coefficient reported by Chaudron is good enough (r=0.66), considering the nature of subjective measurement and the sample size. The raters working on the

Zhang and Halpern project obtained still higher reliability coefficients. It seems that the dependability of the measures has been improving along with sample size and design. Finally, the use *of* statistics is also undergoing refinement, as is evident from the progression starting with the inappropriate matched t-test in Partridge's study to the 2-way analysis of variance in Zhang and Halpern's *study.* All in all, from the methodological point of view, there seems to be a very healthy progression from one project to another.

The discussion above, however has not yet touched on the essential worth of those successive efforts, because the concerns discussed are without exception methodological issues. The important contribution of this line of research, in the opinion of the writer, does not consist in mere refinement *of* methods, but in the specific results they have yielded. Gradually, the research has been moving away from a simple, effective-vs.-ineffective dichotomy towards a more dynamic, interactive understanding which differentiates the intricate relations between causes and outcomes in various dimensions of ESL composing ability. Feedback is no longer understood as a static or mechanical device which either works or fails, but as an organic component *of* the revision process, its effectiveness depending on other components or characteristics of the process, Research interest is shifting from the presence or absence of *the*

effects of a given type of feedback to the question of how effects of feedback vary, depending on other recognizable conditions in an ESL teaching situation. If we compare this dynamic view of feedback with catchy phrases like "writing without teachers" (Elbow 1973), "quantity before quality" (Briere 1966, Erazmus 1960), "precise and immediate (teacher) correction" (Rivers 1978), "systematic (teacher) correction (Rivers 1981), "teaching students to teach each other" (Moffett 1968), "minimal marking" (Hanswell 1983), "PQP" (standing for praise, question and polish, Lyons 1981), to name only a few, we cannot help noticing that the basic understanding of feedback is undergoing a qualitative change. Verification of the dynamic, multi-dimensional effects of feedback is necessary and valuable because it would eventually contribute to our decision to free or not to free ESL composition instruction from fixation upon any quick-and-easy feedback techniques as recommended by various experts.

Summary

This chapter has recounted the transition in composition instruction from the product-centered model to the process-oriented model. In the course of the transition, first the teaching of Ll English composition, then the teaching of ESL composition, have come to recognize the importance of studying the effectiveness of intervention in the revision stage. However, in neither first nor second

language research, have contrastive experiments involving feedback from various sources produced uniform or unequivocal results. Most of the twenty-two studies surveyed and critiqued above have confounded the experimental variable (feedback) with other instructional factors. So the seemingly meaningful results one *way* or another might in a large measure have been due to uncontrolled factors. Nevertheless, in both L1 and *L2* research, there have been interesting results indicating that the effectiveness of a given type of feedback may depend on other factors in the writing process (Beach 1979, Chaudron 1984, Lagana 1974, Zhang and Halpern 1984). Those results tend to substantiate a dynamic and multi-dimensional interpretation of the corrective potential of feedback from a given source. But a great deal more experimentation is needed before the actual workings of corrective feedback can be unraveled.

CHAPTER IV

THE EXPERIMENT

Purpose

The purpose of the present project has much to do with the three ESL studies critiqued in Chapter 111. A careful examination of the three experiments has drawn the present investigator's attention to the following points.

Point 1 -- Partridge found teacher feedback superior to peer feedback. Chaudron found no difference between the two types of feedback. And Zhang and Halpern concluded that the effectiveness of the feedback from a certain source varies with other factors, The somewhat incompatible results, though they can be reconciled on a variety of grounds, clearly point to the necessity of making further inquiries into the issue.

Point 2 -- The first two studies were conducted with very small numbers of subjects (12 to 23). Zhang and Halpern tested 62 subjects. For greater generalizability, the need to further expand the data base must not be overlooked*

Point 3 -- The Partridge study looked at a homogeneous group in terms of ESL competence. Chaudron carried out his experiment with two levels of proficiency considered

separately The possible interaction of level and source of feedback was not analyzed until Zhang and Halpern adopted a 2 x 3 factorial design to partition the observed variance in ESL writing performance according to three identifiable sources: level of proficiency, feedback treatment, and interaction. Zhang and Halpern looked at two proficiency levels: advanced and intermediate. It would be interesting to extend the range to include a still lower section and see whether or not the variance in writing performance would exhibit the same distributive pattern as in Zhang and Halpern's study.

Point 4 -- Because Partridge and Chaudron reported their statistical results upon the basis of total scores, without discriminating differences in subscores, the possibility of different feedback types exerting different effects on different aspects of writing were not adequately clarified. (Chaudron did calculate t-values on the basis of the breakdown of total scores. But the results were omitted from the published report because they closely resembled the overall effects.) Zhang and Halpern broke down the total score into the content/discourse adequacy category and the grammar/mechanics accuracy category, but did not take the trouble to further divide the holistic content/discourse adequacy score into distinct areas such as content, organization and vocabulary. This is another dimension of the issue that deserves careful scrutiny.

Point 5 -- Both Partridge and Chaudron gave out questionnaires to their subjects to solicit their opinions about corrective feedback. Partridge's questions were not worded in such a way that explicit statements of preferences could be elicited. Chaudron surveyed 48 subjects across two proficiency levels and found that ESL learners would like a teacher to read their writings for mistakes and strongly disagree that their writings should be read for mistakes by fellow non-native students. Zhang and Halpern did not include a questionnaire survey in their study. Since the effectiveness of any particular type of feedback is conceivably related to how the recipients feel about the feedback, it is certainly advisable to seek more information about student preferences.

It is with these five points in mind that the present project was conceived and designed. The project represents an attempt to further investigate the different roles of corrective feedback from various sources (Point 1) on a larger scale (Point 2) involving more ESL proficiency levels (Point 3) and more aspects of the overall dependent variable, ESL composing skills (Point 4). At the same time, student preferences were solicited (Point 5) in order to determine whether student choices correspond to the objective statistical analysis of their performances under different experimental conditions. In short, this exploratory study addresses the following research

questions:

1. What sources of corrective feedback are conducive to what aspects of ESL writing improvement at what levels of ESL proficiency?
2. Do ESL learners' preferences for feedback from various sources correspond to the objectively verified effects of *those* feedback types?

Method

Subjects

The subjects were 87 non-native speakers enrolled for the 1985 spring semester in the English Language Institute (ELI) of the University of Hawaii (UH) at Manoa and the English Foundations Program (EFP) of the Hawaii Pacific College (HPC) at Honolulu. In the project, four subjects were eliminated from the experiment because of their incomplete attendance. Two more subjects *at* UH, instead of utilizing feedback for in-class revision as required by the design, copied from draft sheets they had brought to class. *Since* it was impossible to determine what feedback had been incorporated into their out-of-class essays, the two subjects were not counted into the data set. Altogether 81 students completed the experiment. No systematic pattern emerged from an examination of the six uncounted cases. The *6.9%* mortality rate does not seem to suggest any factor that could bias the results of the study.

Among the 81 subjects, thirty-one were females (38.3%), and fifty males (61.7%). 8 subjects originated from Pacific islands (9.9%), 70 from East or Southeast Asia (86.4%), and 3 from Israel, Bulgaria, and Nigeria respectively (altogether 3.7%). Even though the majority of them were from Asia, they represented fairly hetrogeneous L1/dialectal backgrounds (mainland China, Hong Kong, Taiwan, South Korea, Vietnam, Indonesia, Japan, Singapore, Malaysia and the Philippines). The eight Pacific island students came from American Samoa, Palau, the Marshall Islands, Ponape and Truk. The 81 subjects differed widely in length of residence in an English-speaking country, varying from approximately two weeks to fourteen years up to the time of the study. 39 of them (48.1%) had had less than one year's exposure to the naturalistic use of English in an English-speaking country, 42 (51.9%) had experienced over a year's exposure. 63 (77.8%) of them were enrolled at UH, while the remaining 18 (22.2%) were enrolled at HPC.

The subjects were available in three intact groups. 40 of them (49.4%) represented the near total enrollment of three sections of ESL 100 at UH, a 3-credit writing course offered to foreign-born students in lieu of the regular university freshman composition course ENG 100. 23 (28.4%) formed two sections of ELI 73, a non-credit remedial writing course specially designed for international students judged not ready to participate in regular freshman classes. 18 (22.2%) were enrolled in one section of EFP Composition

Level 3 at HPC. The main objective of the course was to reinforce grammar and familiarize the learners with writing-related issues.

Because the 81 subjects were made available through convenience sampling, it was necessary to determine whether the three groups really reflected three normative levels of ESL ability. Most of the foreign students at UH had been placed into ESL 100 or ELI 73 on the basis of a composition-writing exam administered upon their arrival at the university. (For a description of the writing task, see Appendix A.) Their essays were then graded according to the structured ESL Composition Profile devised by Jacobs et al. (1981, see Appendix B). The same test was conducted with the HPC students. Their essays were scored by a rater who had participated in the UH placement test. Then analysis of variance was computed on all the available scores. 15 of the 18-member HPC class took the ELI placement test. Of the 23 ELI 73 subjects, 22 scores were available. But among the 40 ESL 100 subjects, only 18 scores were available. 4 students had applied to UH with such high TOEFL scores that they were exempted from the placement test but they decided to take ESL 100 anyway. 18 students had taken the test one or more than one semester earlier. Failing to reach the minimal score of 70 for enrollment in ESL 100, they all took ELI 73. Upon the successful completion of ELI 73, they were automatically promoted to ESL 100 without taking the same

placement test again. It is reasonable to expect those students with exemption and those students who had completed ELI 73 to be generally more proficient than the average ELI 73 students. The results of the analysis of variance (ANOVA) and the post hoc Student-Newman-Keuls (SNK) multiple-range test are reported respectively in Tables 1 and 2.

Table 1 ANOVA of ELI Placement Test Scores

Source of variance	Sum of squares	Degrees of freedom	Mean square	F
Between	4108	2	2054	73.36*
Within	1477	52	28	
Total	5585	54		

* significant $p < 0.05$

Table 2 SNK Test of ELI Placement Scores

	Upper-Intermediate $\bar{X} = 62.09$	Lower-Intermediate $\bar{X} = 54.93$
Advanced $\bar{X} = 76.56$	14.47 *	21.63 *
Upper-Intermediate $\bar{X} = 62.09$		7.16 *

* significant $p < 0.05$

These results confirm that the three groups represented three distinct sections of the ESL writing proficiency scale, with the ESL 100 students corresponding to the

advanced level, the ELI 73 the upper-intermediate level and the HPC students the lower-intermediate level.

Design

The experiment adopted a 3 x 3 factorial design with type of feedback (teacher, peer, and self) and proficiency level (advanced, upper-intermediate, and lower-intermediate) as independent variables, and 4 impressionistic ratings concerning the informational/rhetorical aspects of ESL writing and 3 objective frequency-count ratings concerning the grammatical/mechanical aspects as the dependent variables. Subjects at each proficiency level were randomly assigned to the three feedback treatments. The distribution of the 81 students is displayed in the following 3 X 3 classification table.

Table 3 Distribution of Subjects by Level and Treatment

		Feedback Treatments			Row Total
		Teacher	Peer	Self	
Proficiency Levels	Advanced	13	14	13	40
	Upper-Intermediate	9	7	7	23
	Lower-Intermediate	6	6	6	18
	Column Total	28	27	18	81

The data were processed with the SPSSX ANOVA program (SPSS 1983) on an IBM 3081 computer at the UH computing center.

Procedure

The feedback treatments were administered during regular class hours over three successive days in February and March 1985.

On Day One, students were told that they would be writing a composition over the next three days. They were also told that they were going to experience different revision procedures in order for their instructors to make an evaluation of the writing program. It was emphasized that the ultimate beneficiaries would be the subjects themselves or future students in the ESL programs. They were expected to take the writing assignment as a regular in-class task so that their performance would not deviate drastically from their normal standards. Then, three topics were put on the blackboard:

1. Compare and contrast mental work with physical labor;
2. Compare and contrast movies and television;
3. Compare and contrast your high school and your college.

The subjects were free to choose any of the topics or suggest their own topics as long as their topics would

involve comparisons or contrast, The researcher then initiated and led a 10-minute casual discussion to ensure that the students understood the meaning of "compare and contrast". Approximately 40 minutes was allocated for the first draft. The subjects were reminded that the purpose of the first session was for them to get their basic ideas down on paper without undue emphasis on linguistic forms. All the drafts were collected at the end of the regular 50-minute session.

One of the more important concerns at this stage was how to control for the difficulty of the assigned content areas. The three topics had been selected in consultation with the regular instructors. None of the topics had been used prior to the experiment, and all the instructors agreed that the topics had relevance to a foreign student's life in the United States. Besides, the topics were broad enough for the subjects to look for some points of interest. The control over rhetorical pattern (comparison/contrast) and the expository nature of the task precluded confounding of topic selection with organizational or stylistic types. Evidence was obtained to the effect that the subjects across the three different levels did not feel any one of the topics significantly more attractive than the others. 20 chose Topic 1 (24.7%), 30 wrote on Topic 2 (37%) and another 30 on Topic 3 (37%). One student decided on a new topic with the researcher's approval (1.2%). He was not included

in the computation of the "goodness of fit" chi-square test (Ferguson 1981:204). The chi-square is non-significant, confirming the null hypothesis that the selection of topics did not exhibit a lop-sided pattern. The subjects across the proficiency levels did not seem to feel much more ease with one topic than another.

This result was substantiated by another test to alleviate a further concern. Normally, poorer students would like to deal with what they felt to be "easier" content areas. Only proficient students try to tackle difficult topics. Therefore, different topic preferences at different levels of competence might suggest degrees of difficulty inherent in assigned topics. A chi-square test of independence (Ferguson 1981:207) was computed on the frequency data displayed in Table 4.

Table 4 Distribution of Topics by Proficiency Levels

Topic Selection	Proficiency Levels: Advanced	Upper-Intermediate	Lower-Intermediate
Topic 1	12	5	3
Topic 2	14	11	5
Topic 3	13	7	10

The chi-square value is non-significant, although the 3-5-10 split in the lower-intermediate group looks

suspicious. Again, there is no basis to assume that any one topic area attracted more subjects than the others. In other words, no topic was perceived to be particularly easy.

Prior to Day Two, students at each level were randomly assigned to three groups to be subjected to feedback from different sources, and the researcher prepared an 18-item checklist (Appendix C) to ensure that the students had a clear idea of what was meant by a comprehensive and balanced evaluation. The 18-item checklist was a synthesis of three checklists currently in use in ELI and six more taken from coursebooks by different authors (Brereton 1978, Clouse 1983, Mattson, Leshing and Levi 1979, Pellegrino 1982, Schoen, Avidson, Gandhi, and Vaugh 1982, Sullivan 1980). Following a pedagogical suggestion by Knapp (1972), all the items were worded as yes/no questions. *Yesn indicates that the requirement of a particular nature has been taken care of in an essay, "No" signifies deficiency in the respect. This checklist was given out to all the subjects. 40 minutes was allowed for revision. The revised texts were collected at the end of the day.

On Day Two, the feedback variable was operationalized in three treatments. The 28 students (34.6%) in the teacher feedback group received their first drafts with teacher corrections. Teacher correction consisted of three forms:

1. underlining mistakes, e.g. "Four years latter, my friends said I had changedn;

2. adding insertion marks where inappropriate omissions occurred, e.g. "You don't have to stand in a line to get ^ ticket*;
3. offering concise comments or suggestions like *The ending is too abrupt", "your remark here contradicts the first sentence of the paragraph".

The students' task on Day Two was to go over all the markings in red and figure out why those markings were there and how to rectify the errors or improve upon the text. Of course, they could also use the checklist to judge their drafts. The researcher was available to answer their questions for further clarification. However, only prompts were given. Outright corrections were never directly provided.

The peer feedback group consisted of 27 students (33.3%), who read the xeroxed copies of one another's drafts. Names were covered up when the copies were being xeroxed. The extra trouble taken to ensure anonymity was intended to encourage candid remarks and straightforward corrections. Peer readers were instructed to check the texts with the 18-item checklist, make explicit corrections, indicate places where they sensed something was amiss and put down whatever comments they felt would facilitate revision. They were allowed to ask the students sitting next to them for help. The whole procedure was planned according to the typical peer correction practices described

by Witbeck (1976). The researcher never helped them with their evaluation.

In the self-feedback (control) group, 26 subjects (32.1%) worked on their own drafts with the guidance of the checklist. They were permitted to consult whatever reference books they wished to, but told not to seek assistance from their classmates.

On Day Three, all the students wrote out the final version in 40 minutes, making as much use as possible of whatever feedback had been provided. It was suggested that they make no drastic change at this stage. The advice was necessary to caution overzealous or embittered subjects against the idea of "making a new start". If that happened, the new draft would not be eligible for analysis because the previous feedback had been thrown away, which would have meant that the time and energy invested in the treatments had been wasted. At the end of Day Three, all the final versions were checked with the drafts. No bold departures were detected. The effects of feedback were unquestionably there in the polished versions.

Measures

The measurement of writing has always presented great uncertainty. So far, the two basic approaches are holistic scoring and frequency-count marking (Cooper 1977). By

'holistic' is meant "any procedure which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing" (Cooper 1977:4). Within the holistic group, some people argue that holistic evaluation should not be guided by any criteria, rubrics, reminders, standards, or structured scales (Lloyd-Jones 1977), while others maintain that holistic, impressionistic evaluation can range from totally unstructured to semi-structured (Cooper 1977). They point out that, even in totally subjective evaluations, readers are following certain rubrics which have been generally agreed upon as essential to the quality of writing. This is the position taken by Jacobs et al. (1981) in devising the ESL Composition Profile. Frequency-count marking relies on tallying elements such as number of errors, total number of words, number of clauses per sentence, number of sentences per composition etc. Each approach has its own strong points and weaknesses* Holistic evaluation gives priority to the communicative function of writing whereas the frequency-count marking tends to treat language as a system independent of meaning. But the objective and methodical frequency count is consistent, while the subjective, holistic judgment is often not. The two approaches have co-existed for a long time. Researchers like Cooper (1977), Evolva, Mamer and Lentz (1980), Jacobs, Zingraf, Wormuth, Hartfiel and Hughey (1981), Kaczmarek (1980), Lloyd-Jones (1977) and Nold and Freeman (1977) share the conviction that

holistic evaluation gets a judge closer to what is essential in writing. On the other hand. Hunt (1965, 1970, 1977), Endicott (1973), Flahive and Snow (19801, Gaies (1980), Witte (1982), and Lim (1983) continue to make a strong case for the use of frequency counts, particularly in the use of T-units. A T-unit is defined as a "minimal terminable unit ▪ minimal as to length, and each would be grammatically capable of being terminated with a capital letter (at one end) and a period (at the other)" (Hunt 1965:21). It is a "single main clause (or independent clause, if you wish) plus whatever other subordinate clauses or non-clauses are attached to, or embedded within, that one main clause" (Hunt 1977:93).

In this study, an eclectic approach allows the two methods to be used simultaneously. The holistic method is used with the content/discourse dimension of ESL writing, and the objective frequency count is used with the grammar/mechanics dimension. Each method is employed to serve the purpose it is generally expected to serve the best. The decision to resort to eclecticism was made not because it appeared to be the easy way out, but because the writer had come to notice some quite serious defects in the "language use* and *mechanics" sections of the ESL Composition Profile. Although it is a reasonably validated scale, a careful match-up of the 18 sample essays and their grammar scores given by 4 model raters as guidance for rater training (Jacobs et al. 1981) revealed interference from

factors that should be judged independent of grammaticality. Compositions with more or less the same grammatical errors-per-T-unit ratio were given conspicuously different scores. And legibility and content seem to be the biggest interferences in assigning grammar scores. The 18 sample essays and the scores serving as guidance for prospective raters were omitted when the same profile was re-published (Hughey, Wormuth, Hartfiel and Jacobs 1983). It was after a careful examination of the 18 sample essays and their scores that the decision was made that the grammatical and mechanical accuracy of ESL writing be measured, instead, by errors-per-T-unit instruments borrowed from Flahive and Snow (1980) and Homburg (1984).

Seven scores are used in the experiment. The content score is determined with reference to 4 descriptors: knowledge, substantiation, development (of thesis), and relevance. Judgments are made with the guidance of 16 criterion questions (Jacobs et al. 1981:92). This category takes up 30 points in a total of 70 (Appendix B).

The organization score synthesizes judgments on 6 descriptors (fluent expression, articulation, succinctness, global structure, logical sequence and cohesion), which are elaborated in 13 criterion questions (Jacobs et al. 1981:93). 20 points are allocated to the category (Appendix B).

The vocabulary score is based upon 4 descriptors: sophistication, effect, derivation and register, explained by 15 criterion questions (Jacobs et al. 1981:94). This category has a maximum of 20 points (Appendix B).

The content/discourse adequacy score is the sum of the three scores above. The maximum score is 70. Since 1981, the authors have slightly modified some of the criterion questions in the categories given above. For more information, see Hughey et al. (1983).

The grammar score is the ratio of the total number *of* grammatical errors to the total number of T-units in a text. It has been empirically proved that, as a single index, the errors/T-unit ratio is not the *best* possible indicator of a syntactically mature or immature writer (Flahive and Snow 1980). However, this experiment is more concerned with how feedback reduces errors than with which quantitative measure is the most accurate or parsimonious representation of the quality of writing. For that purpose, the error ratio seems to have more face validity than other frequency-based instruments, e.g. the mean length of the T-unit (Hunt 1965, 1970, 1977, Lim 1983, Witte 1982), the subordination ratio or clauses/T-unit ratio (Hunt 1965, 1977, O'Donnell, Griffin and Norris 19671, the *complexity* index, also based upon the T-unit (Endicott 1973) or the number of error-free T-units

(Homburg 1984, Larsen-Freeman and Strom 1977, Scott and Tucker 1974).

The same method is used in arriving at a mechanics score, similar to the one used by Homburg (1984). By "mechanics" is meant punctuation, capitalization, paragraphing and spelling. Legibility, one of the 5 descriptors for the mechanics category of the ESL Composition Profile, is omitted in view of the practical difficulty in assigning an objective quantitative score.

The grammar/mechanics accuracy score combines the two scores above, indicating the density of formal irregularities in a given text. Strictly speaking, all the three frequency-based measures are measures of inaccuracy or deficiency, not competence.

Because two fundamentally different approaches are adopted the impressionistic scores and the objective scores in the form of a ratio cannot be added to yield a total score. This might cause some problems in a classroom. But for research purposes, the absence of a total score is not a serious problem. Considering the emphasis on the differential effects of feedback from different sources, it might be a worthwhile loss in return for more reliable and more detailed analysis.

Two native-speaker ELI instructors graded the 81 essays on the first three measures: content, organization, and vocabulary. Both raters had been trained in the use of the scale and had used it for placement and instructional purposes for almost two years. Neither participated in any other aspects of the project. Both rated "blind". One rater scored all the 81 compositions. Then 38 compositions (47%) were randomly picked for the second rater's evaluation. Pearson product-moment correlation coefficients were calculated to test the inter-rater reliability. The results are moderately satisfactory (r=0.72 for content, r=0.66 for organization, r=0.77 for vocabulary, r=0.74 for content/discourse adequacy). For statistical analysis, only the first rater's scores were used.

In the grammar/mechanics category, two other judges counted all the T-units, grammatical errors and mechanical errors. Both read for mistakes without allowing themselves to be distracted by meaning or style. Their frequency counts tallied well. For grammar scores, they reached a satisfactory 81% agreement; for mechanical scores, they obtained 93% agreement. For grammar/mechanics accuracy scores, they reached 87% agreement. The differences were later resolved through discussion and the adjusted scores were used for statistical analysis. For a complete list of all the 567 raw scores, see Appendix D.

Results and Discussion

The means ($\bar{X}$) and standard deviations (S.D.) of the 4 content/discourse scores are displayed in Table 5.

Table 5

Means and Standard Deviations of Content/Discourse Scores

Proficiency	Feedback		Content	Organization	Vocabulary	Content/ Discourse
Advanced	Teacher	$\bar{X}$	22.85	16.23	15.46	54.54
		SD	3.51	2.20	2.26	7.76
	Peer	$\bar{X}$	23.64	16.71	15.71	56.07
		SD	2.90	1.90	1.98	6.44
	Self	$\bar{X}$	20.46	14.46	14.85	49.77
		SD	3.55	2.88	1.34	6.78
Upper-Intermediate	Teacher	$\bar{X}$	19.89	15.44	14.33	49.67
		SD	3.02	2.07	1.66	6.22
	Peer	$\bar{X}$	16.71	12.71	12.86	42.29
		SD	1.89	1.60	2.12	4.99
	Self	$\bar{X}$	20.00	14.71	13.00	47.71
		SD	5.13	3.04	2.83	10.70
Lower-Intermediate	Teacher	$\bar{X}$	19.67	12.50	14.00	44.50
		SD	6.12	2.74	1.67	5.99
	Peer	$\bar{X}$	17.33	13.17	13.67	44.17
		SD	1.51	1.72	0.52	3.37
	Self	$\bar{X}$	17.17	12.50	12.83	42.50
		SD	3.19	2.95	1.94	7.87
Grand Mean			20.43	14.73	14.28	49.44

Table 6
Means of Content/Discourse Scores by Level or Feedback

Category	Group	Content	Organization	Vocabulary	Content/ Discourse
Level	Advanced	22.35	15.82	15.35	53.52
	Upper-Intermediate	18.96	14.39	13.48	46.83
	Lower-Intermediate	18.06	12.72	13.50	43.72
Feedback	Teacher	21.21	15.18	15.41	50.02
	Peer	20.44	14.89	13.65	49.85
	Self	19.58	14.08	13.64	47.54

Figures 1 to 4 provide a graphic display of these data. On the four measures, there is an evident pattern, with the advanced group staying on top of the intermediate level groups. One exception is in the organization category, where upper-intermediate subjects resorting to self-generated feedback achieved a slightly higher group mean than their counterparts at the advanced level. The upper-intermediate group, in its turn, maintained a general performance level higher than that of the lower-intermediate subjects$_I$ except in the peer feedback treatment group, where the lower-intermediate subjects appear to have slightly outperformed the upper-intermediate group. However, the observed differences are actually negligible, considering the standard deviations$_I$ and also the sizes of the scales employed in the study.

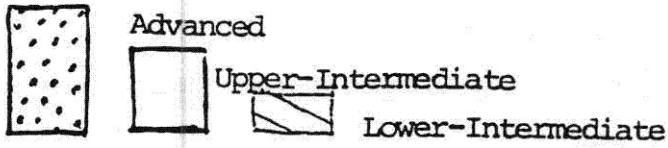


Figure 1

Means of Content Scores Classified by Level and Feedback

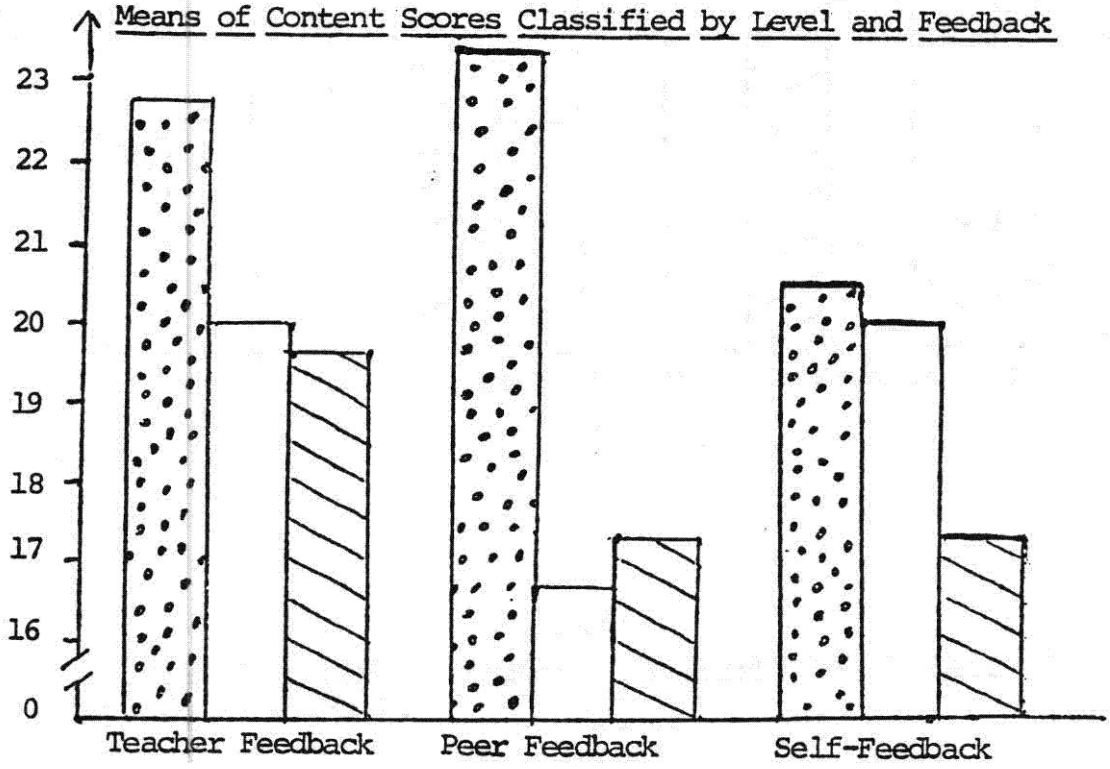


Figure 2

Means of Organization Scores Classified

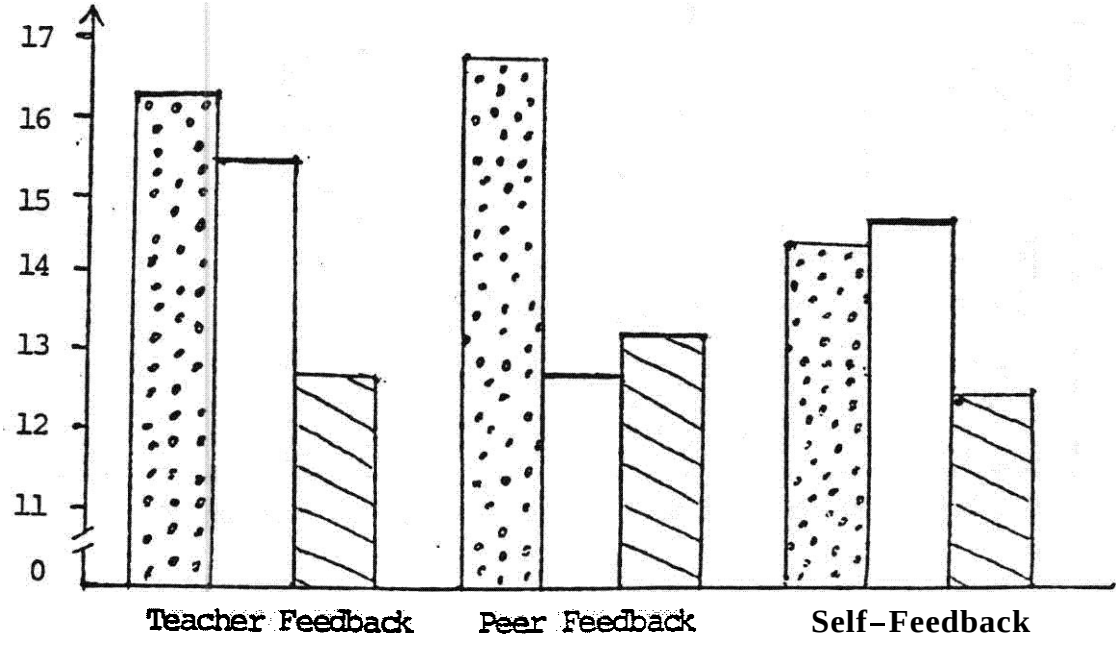


Figure 3

Means of Vocabulary Scores Classified by Level and Feedback

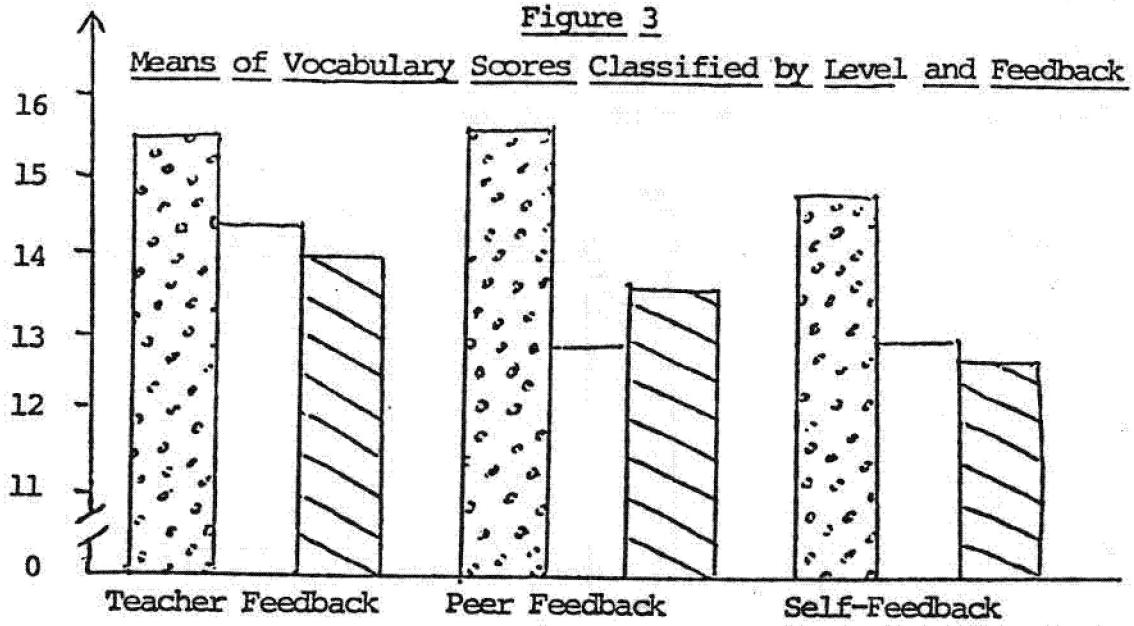


Figure 4

Means of Content/Discourse Adequacy Scores Classified by Level and Feedback

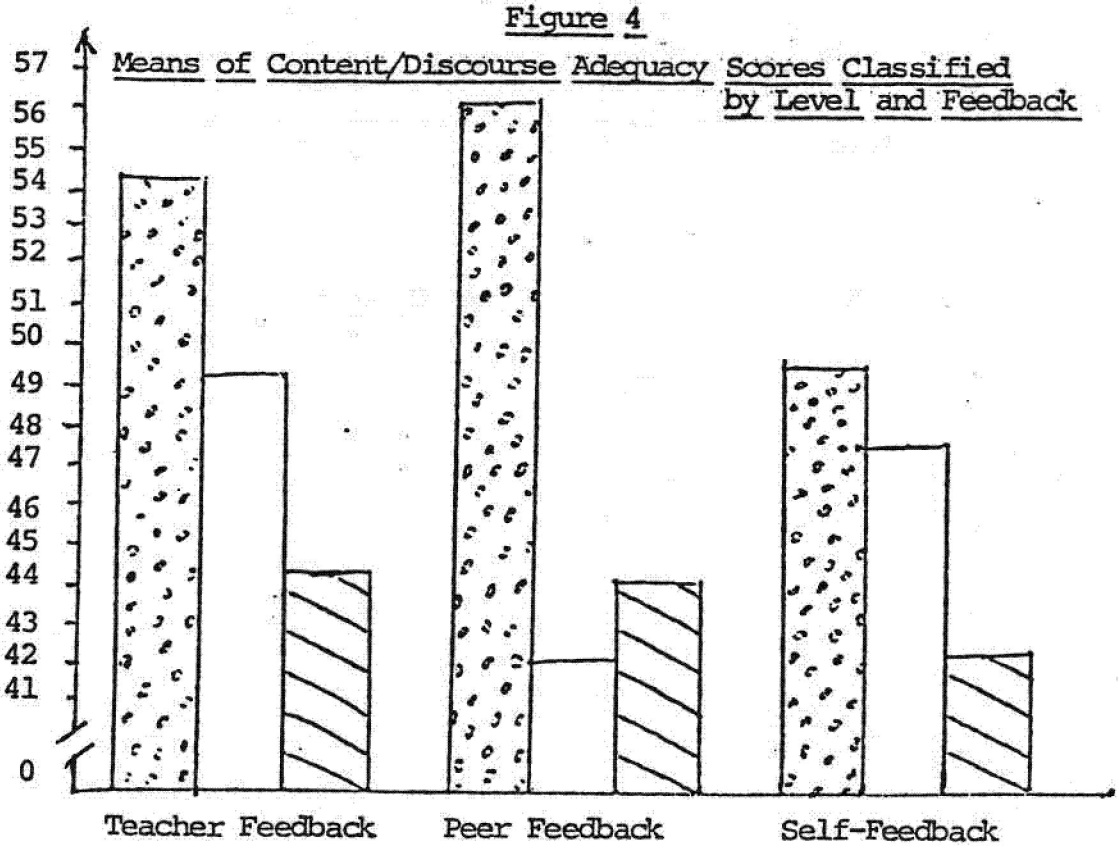


Table 7 summarizes all the means ($\bar{X}$) and standard deviations (S.D.) of the grammar/mechanics scores.

Table 7

Means and Standard Deviations of Grammar/Mechanics Scores

Proficiency	Feedback		Grammar	Mechanics	Grammar/Mechanics
Advanced	Teacher	$\bar{X}$	0.75	0.20	0.88
		SD	0.50	0.13	0.41
	Peer	$\bar{X}$	0.80	0.26	1.06
		SD	0.36	0.19	0.46
	Self	$\bar{X}$	1.07	0.36	1.35
		SD	0.72	0.25	0.89
Upper-Intermediate	Teacher	$\bar{X}$	1.04	0.32	1.36
		SD	0.60	0.36	0.87
	Peer	$\bar{X}$	1.20	0.25	1.46
		SD	0.35	0.14	0.33
	Self	$\bar{X}$	1.11	0.39	1.50
		SD	0.63	0.38	0.88
Lower-Intermediate	Teacher	$\bar{X}$	0.72	0.26	0.98
		SD	0.37	0.20	0.53
	Peer	$\bar{X}$	1.48	0.47	1.94
		SD	0.49	0.34	0.49
	Self	$\bar{X}$	1.76	0.37	2.12
		SD	0.68	0.14	0.71
Grand Mean			1.04	0.31	1.32

Table 8

Means of Grammar/Mechanics Scores by Level or Feedback

Category	Group	Grammar	Mechanics	Grammar/Mechanics
Level	Advanced	Ø.87	Ø.28	1.1Ø
	Upper-Intermediate	1.11	Ø.32	1.43
	Lower-Intermediate	1.32	0.37	1.68
Feedback	Teacher	Ø.84	Ø.25	1.Ø5
	Peer	1.Ø5	Ø.31	1.36
	Self	1.24	Ø.37	1.57

It should be made clear that these means are indices of inaccuracy or deficiency. The higher the score, the lower the performance level. Figures 5 to 7 provide a graphic display of these data.

Figure 5

Means of Grammar Scores Classified by Level and Feedback

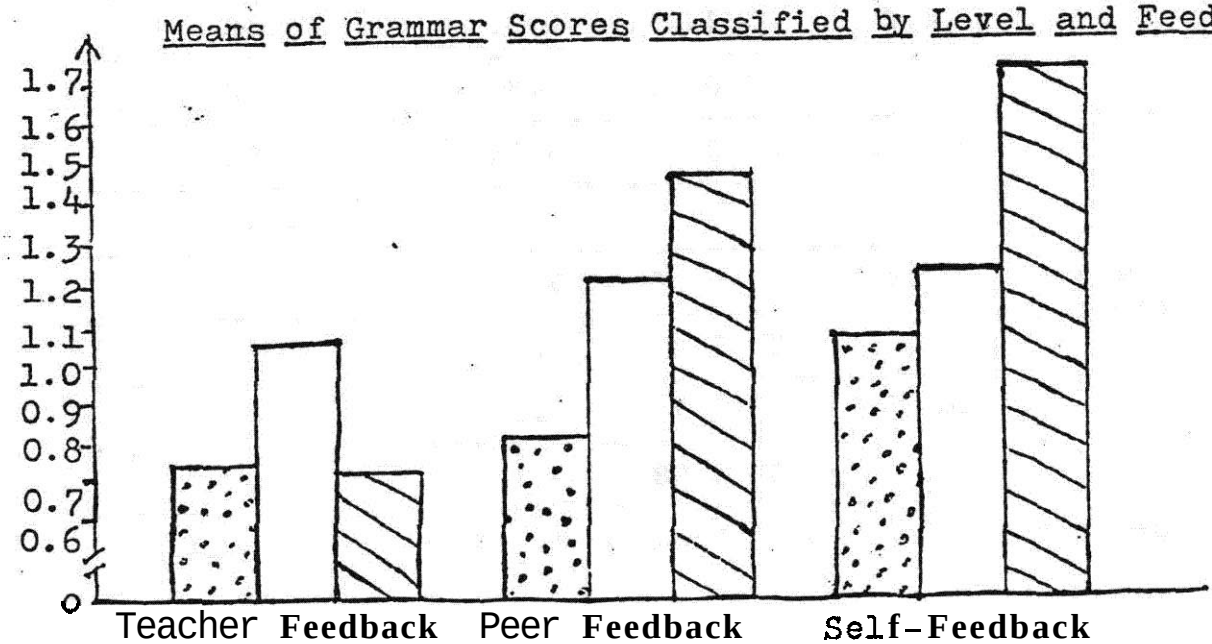


Figure 6

Means of Mechanics Scores Classified by Level and Feedback

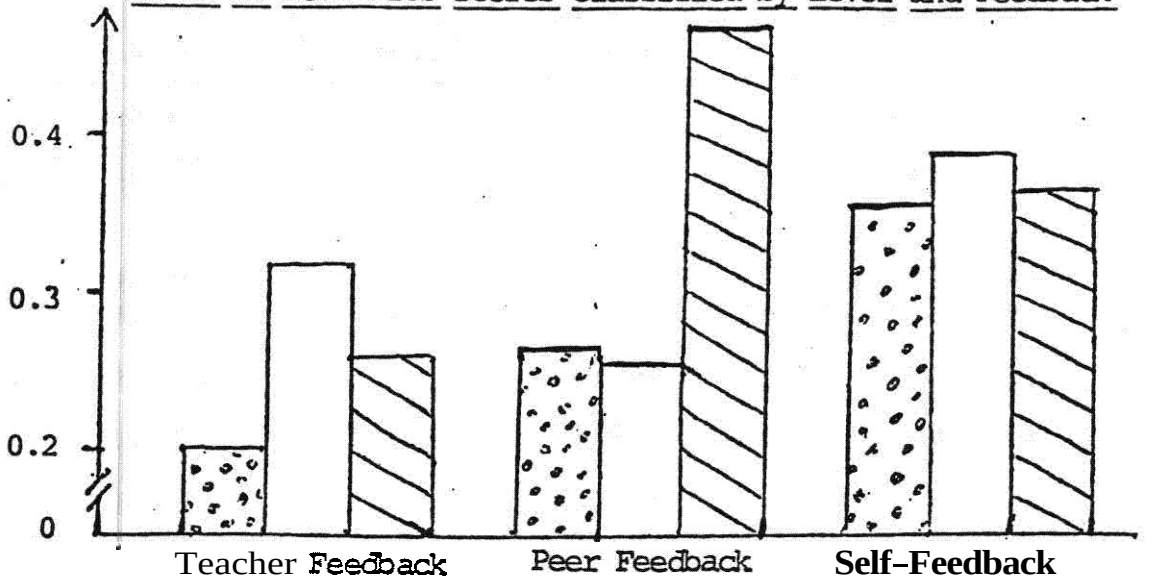
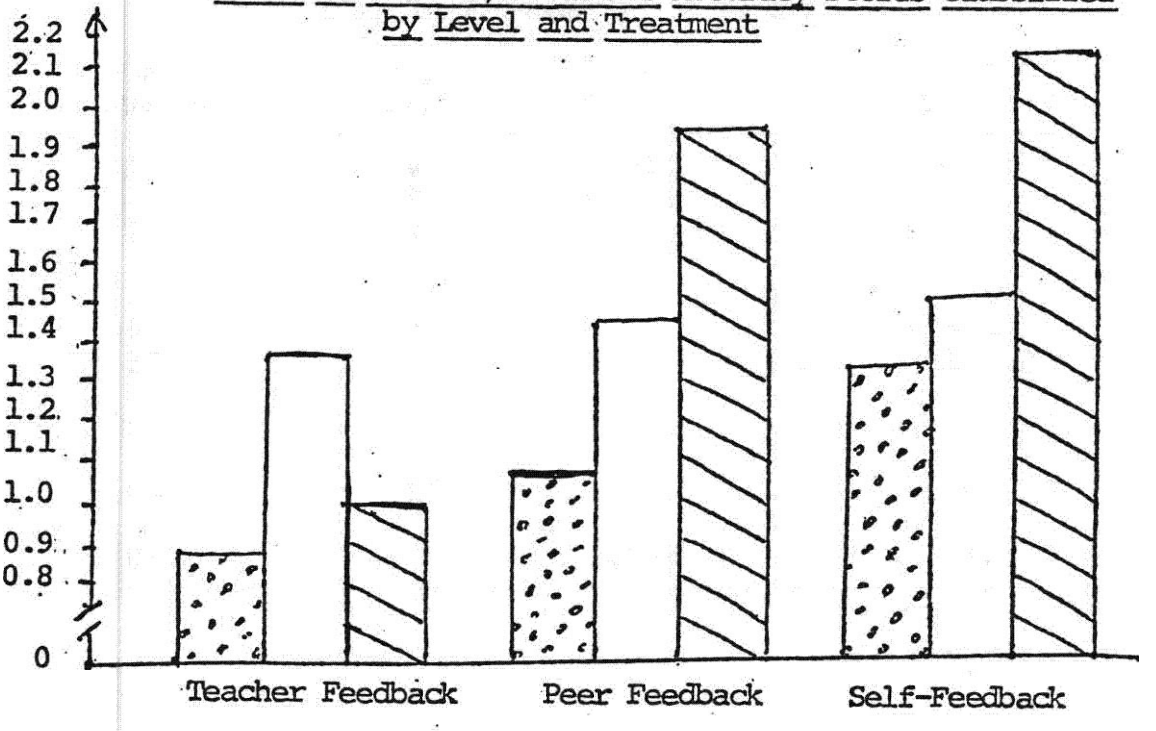


Figure 7

Means of Grammar/Mechanics Accuracy Scores Classified by Level and Treatment



Then separate 2-way ANOVAs were computed on the seven dependent measures. The results are shown in Tables 9 to 15.

Table 9 ANOVA of Content Scores

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	
Main effects	340.18	4	85.04	6.73	*
Level	304.02	2	152.01	12.03	*
Feedback	41.30	2	20.65	1.63	
2-Way Interaction	105.75	4	26.44	2.09	
Level x feedback	105.75	4	26.44	2.09	
Explained	445.92	8	55.74	4.41	*
Residual	909.96	72	12.64		
Total	1,355.88	80	16.95		

* Significant $p < 0.05$

Table 10

ANOVA of Organization Scores

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	
Main effects	140.89	4	35.22	6.26	*
Level	123.49	2	61.75	10.97	*
Feedback	17.73	2	8.87	1.58	
2-Way Interaction	51.82	4	12.96	2.30	
Level x Feedback	51.82	4	12.96	2.30	
Explained	192.72	8	24.09	4.28	*
Residual	405.31	72	5.63		
Total	598.03	80	7.48		

* Significant $p < 0.05$

Table 11

ANOVA of Vocabulary Score

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	
Main effects	83.05	4	20.76	5.65	*
Level	71.60	2	35.80	9.74	*
Feedback	12.83	2	6.42	1.75	
2–way Interaction	7.70	4	1.93	0.52	
Level x Feedback	7.70	4	1.93	0.52	
Explained	90.75	8	11.34	3.08	*
Residual	264.80	72	3.68		
Total	355.56	80	4.44		

* Significant $p < 0.05$

Table 12

ANOVA of Content/Discourse Adequacy Scores

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	
Main effects	1,578.21	4	394.55	8.10	*
Level	1,426.18	2	713.09	14.64	*
Feedback	165.10	2	82.55	1.70	
2–Way Interaction	358.64	4	89.66	1.84	
Level x Feedback	358.64	4	89.66	1.84	
Explained	1,936.84	8	242.11	4.97	*
Residual	3,507.16	72	48.71		
Total	5,444.00	80	68.05		

* Significant $p < 0.05$

Table 13

ANOVA of Grammar Scores

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Main effects	4.90	4	1.22	4.21 *
Level	2.70	2	1.35	4.65 *
Feedback	2.26	2	1.13	3.89 *
2-Way Interaction	2.10	4	0.53	1.81
Level x Feedback	2.10	4	0.53	1.81
Explained	7.00	8	0.88	3.01 *
Residual	20.92	72	0.29	
Total	27.92	80		

* Significant $p < 0.05$

Table 14

ANOVA of Mechanics Scores

Source of Variaton	Sum of Squares	Degrees of Freedom	Mean Square	F
Main effects	0.30	4	0.08	1.27
Level	0.11	2	0.06	0.95
Feedback	0.19	2	0.10	1.63
2-Way Interaction	0.17	4	0.04	0.73
Level x Feedback	0.17	4	0.04	0.73
Explained	0.48	8	0.06	1.00
Residual	4.27	72	0.06	
Total	4.75	80		

Table 15

ANOVA of Grammar/Mechanics Accuracy Scores

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	
Main effects	8.43	4	2.11	4.96	*
Level	4.78	2	2.39	5.62	*
Feedback	3.81	2	1.90	4.48	*
2-Way Interaction	2.36	4	0.59	1.39	
Level x Feedback	2.36	4	0.59	1.39	
Explained	10.79	8	1.35	3.18	*
Residual	30.59	72	0.43		
Total	41.38	80	0.52		

* Significant $p < 0.05$

The first point to be made from the seven ANOVA tables will address the question of whether the proficiency level variable has a main effect, regardless of corrections from different sources. The answer is yes. A highly significant main effect for proficiency level is evident on all dependent measures except the mechanics score. But in the context of the experiment, it is a trivial result because proficiency is defined by measures identical to those used here as dependent variables. A significant F in and of itself provides no new information, although its absence would lead to a question about the adequacy of classification.

Secondly, compared with the proficiency level variable,

the feedback variable dues not appear nearly as important. It has a main effect only on grammar scores and the superordinate grammar/mechanics accuracy scores. This finding agrees with Zhang and Halpern's finding (1984) and also Chaudron's finding (1984). Since it has effects only on one aspect of ESL writing, feedback is not likely to make any dramatic difference in the total score, or overall judgment of a composition. Manipulation of the feedback variable cannot produce as obvious an effect as that of the proficiency level.

Thirdly, none of the interaction effect F's reach the required 0.05 level of significance. That means, if feedback exerts an influence on grammaticality, the effect is more or *less* comparable across the three proficiency levels. Similarly, if it fails to have a measurable impact, as on all the content/discourse scores and the mechanics scores, its lack of effectiveness is also felt to a more or less equal degree on all proficiency levels. This finding conflicts with Zhang and Halpern study (1984), in which a fairly strong interaction effect was identified on the grammatical errors per T-unit score, and the total errors per T-unit score. A plausible explanation is that Zhang and Halpern used two groups differing very little in grammatical competence, as can be seen from the means reported in their study. When *two* experimental groups are very close to each other in pre-treatment proficiency, understandably,

manipulation of feedback becomes crucial. The situation is analogous to holding constant or minimizing the effect of one variable so that the other experimental variable can demonstrate its role to the fullest extent. Under such conditions, it is very probable that a slightly disadvantaged group with favorable input catches up or even surpasses an initially slightly superior group which receives unconducive or even detrimental input. The greater the gap in their pre-treatment proficiency, the less likely it is that the inferior group can outperform the better group in a carefully controlled experiment. Part of the concern in this project is to see whether the variance in writing performance would exhibit the same distributive pattern as that of Zhang and Halpern's study, when the range of proficiency is expanded. That also represents an effort to check how generalizable Zhang and Halpern's discovery of the interaction effect is. With widened differences between the groups, the interaction effect has disappeared.

Post hoc multiple-range tests were conducted following the significant F's. Because the Student-Newman-Keuls test (Ferguson 1983) is a moderate test, not as conservative as the Scheffe test or as lenient as the least significant difference (LSD) test, it was selected for the multiple-range tests.

In the content area, the observed variation is due to different levels of learner proficiency. Manipulation of

feedback produces negligible effects and does not interact with different proficiency levels. The advanced learners proved to be a group significantly different from the other two groups at the 0.05 level. The difference between the upper- and lower-intermediate groups is not statistically significant.

Table 16 SNK Test of Content Scores

	Lower-Intermediate $\bar{X}$=18.06	Upper-Intermediate $\bar{X}$=18.96
Upper-Intermediate $\bar{X}$=18.96	0.90	
Advanced $\bar{X}$=22.35	4.29 •	3.39 *

* Significant p < 0.05

In the organization category, again, only level has a main effect. The gaps between the groups are distinct with the advanced group superceding the upper-intermediate, who in turn supercede the lower-intermediate.

Table 17 SNK Test of Organization Scores

	Lower-Intermediate $\bar{X}$=12.72	Upper-Intermediate $\bar{X}$=14.39
Upper-Intermediate $\bar{X}$=14.39	1.67 *	
Advanced $\bar{X}$=15.83	3.11 •	1.44 *

* Significant p < 0.05

In the vocabulary area, the picture is somewhat different. The mean of the lower-intermediate students is slightly higher than that of the upper-intermediate students, but the difference is negligible. Advanced students, however, are clearly superior to the others in the use of vocabulary.

Table 18 SNK Test of Vocabulary Scores

	Upper-Intermediate $\bar{X}$=13.48	Lower-Intermediate $\bar{X}$=13.50
Lower-Intermediate $\bar{X}$=13.50	0.02	
Advanced $\bar{X}$=15.35	1.87 *	1.85 *

* Significant $p < 0.05$

The composite content/discourse adequacy scores retain the pattern of the content scores and the vocabulary scores. The advanced students are significantly better than the two intermediate groups. But the upper-intermediate learners are not significantly better than the lower-intermediate ones.

Table 19 SNK Test of Content/Discourse Adequacy Scores

	Lower-Intermediate $\bar{X}$=43.72	Upper-Intermediate $\bar{X}$=46.83
Upper-Intermediate $\bar{X}$=46.83	3.11	
Advanced $\bar{X}$=53.53	9.81 *	6.70 *

* Significant $p < 0.05$

In the area of grammar, both proficiency level and feedback have a main effect, but no interaction effect is found. First, crossing over the feedback treatments, the three proficiency levels are arranged according to the magnitude of the means for a multiple-range test. The advanced students have the lowest mean and prove to be significantly different from the lower-intermediate group. The difference between the upper- and lower-intermediate learners is not significant. Crossing over the levels, teacher feedback proves to be definitely more useful than self-feedback only.

Table 20

SNK Test of Grammar Scores across Feedback

	Advanced $\bar{X}$=0.87	Upper-Intermediate $\bar{X}$=1.11
Upper-Intermediate $\bar{X}$=1.1	0.24	
Lower-Intermediate $\bar{X}$=1.32	0.45 *	0.21

* Significant $p < 0.05$

Table 21

SNK Test of Grammar Scores across Proficiency Levels

	Teacher Feedback $\bar{X}$=0.84	Peer Feedback $\bar{X}$=1.05
Peer Feedback $\bar{X}$=1.05	0.21	
Self-Feedback $\bar{X}$=1.24	0.40 *	0.19

* Significant $p < 0.05$

In the superordinate grammar/mechanics accuracy scores, the same pattern is retained with proficiency level or feedback type disregarded. This is not at all surprising, considering the nature of the composite score and the small fluctuations on the mechanics measure. The correlation

found to be as high as 0.91. The advanced learners are clearly more accurate than lower-intermediate students, and the teacher feedback superior to self-feedback in dealing with grammatical inaccuracies.

Table 22

SNK Test of Grammar/Mechanics Accuracy Scores Across Feedback Treatments

	Advanced $\bar{X}$=1.10	Upper-Intermediate $\bar{X}$=1.43
Upper-Intermediate $\bar{X}$=1.43	0.33	
Lower-Intermediate $\bar{X}$=1.68	0.58 *	0.25

* Significant $p < 0.05$

Table 23

SNK Test of Grammar/Mechanics Accuracy Scores Across Proficiency Levels

	Teacher Feedback $\bar{X}$=1.05	Peer Feedback $\bar{X}$=1.36
Peer Feedback $\bar{X}$=1.36	0.31	
Self-Feedback $\bar{X}$=1.57	0.52 *	0.21

* Significant $p < 0.05$

In summary, the basic findings of the experiment are:

1. The informational/rhetorical aspects of ESL writing proficiency are not likely to be influenced by the manupulation of feedback sources.
2. Source of feedback influences the grammaticality of a revised ESL composition.
3. Teacher feedback has been empirically proved more helpful than self-feedback in dealing with grammatical deficiencies.
4. At the three levels examined in the study, it is not certain whether teacher feedback is more beneficial than peer feedback. Also uncertain is the effect of peer feedback as compared with that of self-feedback. Post hoc comparisons do not support the claim that ESL learners left alone to figure out how their writings can be improved can make more progress than comparable students with feedback from teachers or peers.
5. The main effect of feedback does not interact with the proficiency variable at a significant level.

The absence of an interaction effect precludes a posteriori comparisons of the effects of feedback on separate proficiency levels considered in isolation. But in studying the interaction effect in Zhang and Halpern's study (1984) and the absence of such an effect in the present experiment, a careful researcher might discern clues to

warrant future hypotheses concerning the optimal range of proficiency where manipulation of feedback sources might produce maximal positive effects. What follows is no longer what can be called "results", but some conjectures leading to future research. The tables of means and the histograms tentatively indicate that advanced ESL learners are not sensitive to the manipulation of feedback. Upper-intermediate students also seem to be capable of worthwhile self-evaluation and editing. But the lower-intermediate learners appear to be quite sensitive to feedback types, especially on the grammar score. And teacher feedback is apparently more effective with lower level learners than peer feedback, which, in turn, appears more helpful than self-feedback. It seems that, once a learner's overall ESL ability has reached a certain point, alternative feedback treatments have little influence on his or her written production. It *is* those not so proficient learners who might respond differentially to feedback from various sources. From the results we have now, traditional teacher feedback still seems to be a very promising practice. But those conjectures must await further experimentation.

Summary

With a population of ESL learners as are represented by the subjects in this study, choice among teacher feedback, peer feedback and self-feedback as a means of facilitating

ESL composition revision does not seem to make any significant difference to the informational/rhetorical aspects of composing skills. The only aspect where manipulation of feedback sources causes significant variation is grammaticality, which, understandably, is not a small concern in an ESL writing program. Teacher input definitely stimulates better syntactical revision than self-generated feedback$_I$ and there is tentative evidence to the effect that peer feedback may also be superior to self-feedback in reducing grammatical irregularities. Although no interaction effect was discovered in the experiment, there has been evidence suggesting that teacher feedback might work better with lower level learners than with higher level learners. Besides, the obtained levels of significance for the 2-way interaction in the content and organization areas (0.09 and 0.07 respectively) are close enough to the required 0.05 level to stimulate more research to confirm, or modify, the findings of this study.

CHAPTER V

THE SURVEY

Purpose and Method

On Day Three of the project, each subject was asked to answer two questions:

1. If you are given a choice between teacher evaluation and non-teacher evaluation before you write the final version, which would you prefer?
2. If you are given a choice between peer evaluation and self-evaluation before you write the final version, which would you prefer?

The researcher explained that answers to these questions would help to fit classroom procedures to students[1] preferences so that their initiative and co-operation could be better mobilized. The students were instructed not to think about only what they had expeienced in this particular project. It was their general attitude towards various feedback types rather than the specific techniques used in the project that the researcher was interested in. The subjects were assured that their individual preferences would not be disclosed to their regular instructors, no matter what corrective routines the instructors had been using.

Many books and journal articles have been written about

the affective advantage of non-teacher feedback. But the bulk of the literature is based upon presumptions rather than research. Chaudron (1984) sums up the affective advantage of peer feedback in the following words:

> 2. Peers' feedback is more at the learner's level of development or interest, thus perceived as more relevant than the superior or old teacher's feedback;
> 3. Since multiple peers may be used, learners gain a sense of a wider audience than the one teacher;
> 4. Learners' attitudes toward writing can be enhanced by the more socially supportive peers.

But these claims have not yet been objectively verified. For example, do students actually feel that their teachers are "nit-pickers" (Moffett 1968:195) while their peers are an "immediate, socially appropriate audience" providing a "more compelling impetus" for the student writer to revise (Clifford 1981:50)? Do they actually give priority to peer comments over teacher judgment (Pierson 1967)? Do they sense more social support in peer feedback than in teacher guidance (Elbow 1973)? Do they actually find peer input more "relevant" (Chaudron 19841, more comprehensible than teacher commentary? Hardly any research results are available to verify the alleged appeal of non-teacher corrective procedures. As a matterof fact, such advantages of peer feedback in particular are admittedly "assumed

advantages without "formal support" (Witbeck 1973). In the 22 studies surveyed in Chapter III, only 4 have reported findings relevant to this question. Pfeiffer (1981) found that peer feedback did not lower L1 students' anxiety on a writing assignment. Maize (1954) reported that L1 students exposed to peer correction or teacher correction felt the same about their writing tasks and their writing instructors. Partridge (1981) noted that ESL learners doubted the quality of peer input at the revision stage, and Chaudron (1984) confirmed the same feeling in his questionnaire results. None of the reported empirical findings support the alleged affective advantage of non-teacher feedback. Whatever affective evidence exists in favor of non-conventional, non-teacher feedback is at best anecdotal. Consequently its generalizability is tenuous. The survey reported below is a partial replication of Chaudron's questionnaire survey with the purpose of eliciting unambiguous statements of preference. These answers serve as a basis for a solution to the second research question, i.e. do ESL learners' preferences correspond to the objectively verified effects of the feedback from the three different sources?

Results and Discussion

In answering the first question, seventy-six (93.8%) of the 81 subjects chose the traditional teacher feedback over

non-teacher feedback. Three (3.7%) preferred non-teacher corrective feedback. Two (2.5%) did not answer the question. A one-way chi-square test was conducted with Yates correction {Hatch and Farhady 1982). The result is highly significant ($\chi^2 = 65.6$, $p < 0.001$). On Question 2, forty-nine (60.5%) stated preference for peer feedback, twenty-eight (34.6%) for self-generated correction. Four students (4.9%) failed to make a choice. The one-way chi-square result is also significant ($\chi^2 = 5.2$, $p < 0.05$). These results *support the* position that ESL learners as a whole welcome teaches correction. It is also obvious from the answers that their preference for peer feedback is contingent upon the unavailability of teacher guidance. In other words, peer help is seen as the second best thing preferable only when the teacher *stops* doing correction or evaluation, which happens frequently in ESL writing classes. This finding, along with Partridge's and Chaudron's findings, refutes speculations to the effect that students write off the comments by *a* teacher *by* saying, "Adults just can't understand", or "English teachers are nit-pickers anyway" (Moffett 1968:195). This type of presumption needs to be scrutinized very carefully because so much of the student-centered writing theory assumes the intrinsic unpopularity of teacher feedback. It is still *too* early *to* conclude whether this premise is right or wrong, but the evidence from the three independent ESL studies over four years has converged on the same contrary conclusion. If so verified by future research, ESL teachers and

researchers would have to prepare themselves for yet another "paradigm shift".

From the statistics, it is irrefutable that there are real, substantial differences in the type of feedback chosen by ESL learners. But are *there* other conditions working on their choices of corrective feedback? For example, is it possible that feedback selection is associated with proficiency? *With* regard to the teacher vs. non-teacher choice, with the *2* subjects who did not answer the question as missing cases, seventy-six *(96.2%)* of the 79 surveyed chose teacher correction. Such a high percentage shows that ESL learners are predominantly in favor of teacher correction, regardless of differences in proficiency. And the same can be said *of* other conditions such as differences *in* sex, length of residence in an English-speaking country or ethnicity. In the *case* of ethnicity, it should be mentioned that the sample is typical only of the population *of* ESL learners in Hawaii. Seventy subjects (86.4%) were Orientals, eight (9.9%) were Pacific islanders. The subjects' almost unanimous preference for teacher feedback, interestingly, is not adequately supported by the results of the experiment. The reader will recall that, on most of the dependent measures, the feedback variable has no main efect. On the grammar score, which is conceivably a big concern on ESL learners' minds, teacher feedback is definitely *more* effective than self-feedback, but *it* does not have a

statistically significant superiority over peer feedback. An inspection of the table of means will show that teacher feedback is generally more helpful than peer feedback, but not significantly so in a statistical sense. Here, discretion is advisable before a decision is made as to whether to follow the students' preferences or the results of the experiment. But there is no obvious reason not to play safe by employing teacher feedback.

With regard to the peer vs. self choice, objective results again produce no statistically significant differences between the two, but descriptive statistics also suggest that peer feedback may be more helpful than self-feedback. The subjects* preferences correspond to the table of means but are not convincingly supported by the SNK multiple-range test results. Like their preference for teacher feedback, their general preference for peer feedback is shown by chi-square results to be independent of such factors as proficiency, sex, ethnicity or length of residence in an English-speaking country. Tables 24 to 27 are contingency tables showing relationships between choice of feedback and other conditions of the sample.

Table 24

Contingency Table of Relationship

Between Feedback Selection and Proficiency Level

	Peer Feedback	Self-Feedback	Total
Advanced	25	12	37
Upper-Intermediate	12	11	23
Lower-Intermediate	12	6	18
Total	49	29	78

$\chi^2 = 1.59$ non-significant

Table 25

Contingency Table of Relationship

Between Feedback Selection and Sex

	Peer Feedback	Self-Feedback	Total
Male	28	20	48
Female	21	9	30
Total	49	29	78

$\chi^2 = 0.63$ non-significant

Table 26

Contingency Table of Relationship Between Feedback Selection and Ethnicity

	Peer Feedback	Self-Feedback	Total
Orientals	42	25	67
Pacific Islanders	5	3	8
Total	47	28	75

$\chi^a = 0$ non-significant

Table 27

Contingency Table of RElationship Between Feedback Selection and Length of Residence in an English-speaking Country

	Peer Feedback	Self-Feedback	Total
Over a year	24	16	40
Under a year	25	13	38
Total	49	29	78

$\chi^2 = 0.09$ non-significant

To sum up, survey results indicate that ESL learners at or above the intermediate level almost unanimously prefer conventional teacher feedback to any other types of feedback. From their voluntary answers, no basis is found for the claim that teacher correction is intrinsically unpopular with ESL learners. When teacher feedback is not

available, ESL learners turn to peers for clues to revision and try to avoid self-feedback. Their predilection for teacher feedback, however, does not seem to be based upon an objective judgment of the comparative effectiveness of the feedback from different sources. Their obvious lack of enthusiasm for self-feedback, on the other hand, agrees well with the result of the experiment. As a comprehensive picture, their preferences seem to cross over factors like ESL proficiency, sex, ethnicity, and familiarity with the natural use of English. At present, there is some reason to be suspicious of a currently quite prevalent claim that non-teacher feedback in a shared-authority educational setting has more appeal to learners than the orthodox teacher feedback. At least in the ESL situation, teacher feedback is more enthusiastically sought than its recent rivals, peer feedback and self-feedback.

CHAPTER VI

CONCLUSION

Basic Findings

The influence of corrective feedback from various sources was shown to vary, depending on the aspects of ESL writing proficiency being measured. No significant differences were found in the informational and/or rhetorical dimensions. Nor is there a significant difference due to feedback treatment in mechanical accuracy, but the level of grammaticality, expressed as an errors/T-unit ratio was found to be sensitive to the manipulation of the sources of feedback. Teacher evaluation is definitely superior to self-feedback as stimulation for successful grammatical revision, and might even have some advantage over peer feedback for the same purpose. With the tentative evidence available from the project, this researcher is willing to posit that teacher and peer feedback is more beneficial to lower level learners than self-feedback, and that conventional teacher feedback, when provided as intervention in the revision process, may very well hold its own against the more recently advocated peer feedback.

Survey results indicate that ESL learners predominantly prefer teacher feedback to peer feedback and try to avoid self-feedback. No convincing empirical evidence is

available from the experiment to justify their almost unanimous predilection for teacher feedback. But their perception of self-feedback as ineffective in dealing with grammaticality agrees well with the pattern in which grammar scores vary under the experimental conditions. Student choices of feedback from various sources are particularly interesting because they point to a weak link in the logical reasoning underlying the advocacy of the innovative non-teacher corrective feedback. If ESL learners genuinely welcome teacher judgment and teachers have proved to be as effective as, if nut more so than, peer readers, as can be seen from this experiment, should not we reconsider some of the accusations levelled at *the* teacher? That, of course, does nut imply that writing instructors should "debunk" peer feedback, The positive potential of peer feedback is surely there. But do we know enough about peer feedback to justify the replacement of teacher feedback by peer feedback?

Some Problems

This study suffers from some deficiencies. One has to do with the lack of information about any longitudinal or carry-over effects of respective treatments. No reliable claims about the potential long-term benefits or disadvantages of any corrective treatment can be formed on the basis of a single "one-shot" study like this one. Another issue concerns the limited numbers of subjects in

the nine cells of the 3 x 3 factorial design. It is possible that random assignment on such a small scale could not completely counter-balance pre-existing differences. Besides, the measures employed in the experiment are far from ideal. While acceptable, the inter-rater correlations of those holistic scores were not as high as they were expected to be. And the objective frequency-count measures tend to blur qualitative differences in the kinds of errors made. For example, a minor slip is counted as much as a serious error obstructing communication. If better instrumentation, impressionistic or objective, could be devised in the future, the present results might be found to be inadequate or even incorrect. Finally, it is only appropriate to add that nothing in this study distinguishes between "good" and "bad" revisers. The two words, "good" and "bad" are convenient labels to represent cognitive differences. The experiment is based upon the assumption that all the learners are endowed with a comparable level of sensitivity or susceptibility to corrective feedback. Therefore, variation in their writing performance is directly linked with feedback treatments. It is quite conceivable that good revisers, even operating within the constraints all non-native speakers must face, utilize feedback from whatever sources in a manner different from a bad reviser. Under identical circumstances, good revisers may make far more progress than bad revisers.

In conclusion, future research is necessary to test

whether the caveats above are justified or not. Pending better executed inquiries, present results caution ESL writing instructors against an oversimplified notion of the efficacy or affective advantage of any individual source of feedback. A strong ESL writing program will have to identity specific needs at specific levels and supply appropriate corrective procedures. It is quite misleading to assert the superiority of any one type *of* corrective feedback. Not all aspects *of* ESL writing proficiency or all levels of ESL learners are likely to benefit maximally from identical corrective feedback, However, the findings of this work need to be replicated and elaborated *by* future research.

APPENDICES

Appendix A

ELI Placement Test

INSTRUCTIONS: Your job is to write a composition on one of the topics below.

1. Select one of the topics. Do not write on all of them.
2. Begin writing as soon as you have selected a topic.
3. Write on one side of a page only.
4. Write on every other line.
5. Plan your writing for approximately 40 minutes.
6. You may make an outline, or write a draft first if you wish. Simply draw a large X through the parts you do not want the instructor to read.
7. Write your name (family name first), date and the number of the topic at the top of your sheet.

TOPICS:

1. In an automated society of the future people may have a lot of leisure time. What would be the advantages and disadvantages of this?
2. Discuss the points for and against euthanasia (mercy-killing) and the circumstances, if any, under which it is justified.
3. Something I've changed my mind about.
4. What are three of the greatest areas of contrast between life in your country and the U.S., as you have seen it so far?

Appendix B

ESL COMPOSITION PROFILE

STUDENT DATE TOPIC

	SCORE	LEVEL	CRITERIA	COMMENTS
CONTENT		30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic	
		26-22	GOOD TO AVERAGE some knowledge of subject • adequate range • limited development of thesis. mostly relevant to topic, but lacks detail	
		21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic	
		16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate	
ORGANIZATION		20-18	EXCELLENT TO VERY GOOD: fluent expression a idea dearly stated/supported • succinct • well-organized • logical sequencing • cohesive	
		17-14	GOOD TO AVERAGE somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing	
		13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development	
		9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate	
VOCABULARY		20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register	
		17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage *but meaning not obscured*	
		13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • *meaning confused or obscured*	
		9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate	
LANGUAGE USE		25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions	
		21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions *but meaning seldom obscured*	
		17-11	FAIR TO POOR: major problems *in* simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • *meaning confused or obscured*	
		10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate	
MECHANICS		5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing	
		4	GOOD TO AVERAGE occasional errors of spelling, punctuation, capitalization, paragraphing *but meaning not obscured*	
		3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • *meaning confused or obscured*	
		2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate	

TOTAL SCORE READER COMMENTS

Appendix C

A CHECKLIST

1) Do you find any particularly interesting ideas in the essay?
2) Are there sufficient details to support those prominent ideas in the essay?
3) Do you find any instances of unnecessary repetition?
4) Is there a consistent point of view maintained throughout the essay?
5) If there is a change in the writer's point of view, is it justified?
6) Does the essay have a definite point or points to make?
7) Can you locate the thesis statement?
8) Do you see any particular order in the development of the essay?
9) Does the order of the paragraphs reflect distinct stages in the development?
10) Do all paragraphs have clear indications of what they are about?
11) Within each paragraph, are the sentences varied and logically connected?
12) Is there a proper ending to the essay?
13) Are there any instances where the words actually used are obviously not the words the writer intends to use?
14) Do you feel the words used are appropriate for this type of writing?

15) Do you find a reasonable number of synonyms or antonyms in the comparisons and contrasts made?

16) Do you find any particularly clever or effective combination of words?

17) Have you proofread for grammatical errors? (e.g. Subject-verb agreement, noun-pronoun agreement, singular/plural distinction in the ending of nouns, specific verb forms, consistency in tense, sentence completion, articles, etc.)

18) Have you proofread for mechanical accuracy? (e.g. spelling, capitalization, punctuation, indentation, etc.)

Appendix D

Subject Number	Proficiency Level	Feedback Treatment	Topic Choice	Content Score	Organization Score	Vocabulary Score	Content/ Discourse Adequacy Score	Grammar Score	Mechanics Score	Grammar/ Mechanics Accuracy Score
1	Advanced	Teacher	3	18	13	12	43	Ø.59	Ø.36	Ø.95
2	"	"	1	21	15	14	50	Ø.45	Ø.50	Ø.95
3	"	"	2	27	18	18	63	Ø.64	Ø.09	Ø.73
4	"	"	3	22	16	17	55	Ø.90	Ø.24	1.14
5	"	"	1	24	17	16	57	1.00	Ø.15	1.15
6	"	"	2	20	13	13	46	1.41	Ø.27	1.68
7	"	"	2	28	20	18	66	Ø.17	Ø.05	Ø.22
8	"	"	3	22	17	15	54	Ø.21	Ø.07	Ø.28
9	"	"	3	27	18	18	63	Ø.38	Ø.10	Ø.48
10	"	"	2	27	18	18	63	Ø.45	Ø.12	Ø.57
11	"	"	2	18	13	13	44	Ø.70	Ø.20	Ø.90
12	"	"	2	20	16	13	49	Ø.96	Ø.21	1.17
13	"	"	3	23	17	16	56	Ø.94	Ø.26	1.20
14	"	Peer	1	27	18	18	63	Ø.38	Ø.07	Ø.45
15	"	"	3	26	18	17	61	Ø.49	Ø.16	Ø.65
16	"	"	1	24	18	17	59	Ø.50	Ø.37	Ø.87
17	"	"	1	21	13	13	47	Ø.56	Ø.28	Ø.84
18	"	"	3	20	15	13	48	1.00	Ø.25	1.25
19	"	"	2	21	15	14	50	1.00	Ø.64	1.64
20	"	"	3	22	18	15	55	1.22	Ø.41	1.63
21	"	"	3	26	18	18	62	1.29	Ø.29	1.58
22	"	"	2	21	15	13	49	1.27	Ø.04	1.31
23	"	"	3	27	19	18	64	Ø.73	Ø.09	Ø.82
24	"	"	1	22	17	17	56	Ø.38	Ø.15	Ø.53

Subject Number	Proficiency Level	Feedback Treatment	Topic Choice	Content Score	Organization Scot.	Vocabulary Score	Content/ Discourse Adequacy Score	Grammar Score	Mechanics Score	Grammar/ Mechanics Accuracy Score
25	"	"	1	27	18	17	62	0.89	0.13	1.02
26	"	"	3	20	14	14	48	1.11	0.60	1.71
27	"	"	1	27	18	16	61	0.31	0.22	0.53
28	"	Self	2	24	15	15	54	0.30	0.08	0.38
29	"	"	2	27	19	17	63	0.35	0.02	0.37
30	"	"	1	20	15	14	49	0.59	0.56	1.15
31	"	"	2	24	17	15	56	0.51	0.29	0.80
32	"	"	2	22	18	15	55	0.41	0.07	0.48
33	"	"	1	17	13	12	42	0.57	0.11	0.68
34	"	"	3	21	15	14	50	0.70	0.33	1.03
35	"	"	1	19	13	15	47	0.73	0.58	1.31
36	"	"	3	16	10	14	40	1.21	0.28	1.49
37	"	"	3	14	9	17	40	1.44	0.73	2.17
38	"	"		22	16	16	54	1.64	0.36	2.00
39	"	"	2	21	15	15	51	1.76	0.78	2.54
40	"	"	1	19	13	14	46	2.68	0.50	3.18
41	Upper-Intermediate	Teacher	2	20	18	14	52	0.18	0.10	0.28
42	"	"	3	19	17	14	50	0.90	0.05	0.95
43	"	"	2	23	17	17	57	1.41	1.14	2.55
44	"	"	1	21	15	16	52	2.27	0.64	2.91
45	"	"	2	26	18	16	60	0.70	0.04	0.74
46	"	"	2	17	13	12	42	0.88	0.13	1.01
47	"	"	3	18	14	14	46	0.58	0.25	0.83
48	"	"	2	18	14	13	45	1.12	0.35	1.47
49	"	"	2	17	13	13	43	1.30	0.22	1.52
50	"	Peer	2	17	13	15	45	0.69	0.39	1.00

Subject Number	Proficiency Level	Feedback Treatment	Topic Choice	Content Score	Organization Score	Vocabulary Score	Content/ Discourse Adequacy Score	Grammar Score	Mechanics Scare	Grammar/ Mechanics Accuracy Score
51	"	"	2	19	13	13	45	1.03	0.18	1.21
52	"	"	1	17	15	13	45	0.97	0.30	1.27
53	"	"	1	15	12	12	39	1.26	0.48	1.74
54	"	"	1	19	14	16	49	1.24	0.24	1.48
55	"	"	3	14	10	11	35	1.74	0.24	1.98
56	"	"	3	16	12	10	38	1.48	0.03	1.51
57	"	Self	3	17	14	13	44	0.56	0.09	0.65
58	"	"	2	22	17	15	54	0.77	0.31	1.08
59	"	"	2	14	9	9	32	1.31	1.17	2.48
60	"	"	1	16	14	11	41	2.30	0.53	2.83
61	"	"	2	18	14	11	43	1.42	0.33	1.76
62	"	"	3	25	17	15	57	0.95	0.11	1.06
63	"	"	3	28	18	17	63	0.47	0.19	0.66
64	Lower Intermediate	Teacher	3	21	17	14	51	0.20	0.11	0.31
65	"	"	1	18	12	15	45	0.52	0.19	0.71
66	"	"	2	18	14	12	44	0.56	0.00	0.56
67	"	"	1	17	12	15	44	0.91	0.26	1.17
68	"	"	3	13	9	12	34	1.20	0.50	1.70
69	"	"	2	21	11	16	48	0.90	0.48	1.38
70	"	Peer	3	19	13	14	46	1.13	0.92	2.05
71	"	"	3	19	15	14	48	1.68	0.74	2.42
72	"	"	1	17	14	13	44	1.97	0.28	2.25
73	"	"	2	17	13	14	44	1.96	0.32	2.28
74	"	"	3	17	14	14	45	1.36	0.00	1.36
75	"	"	2	15	10	13	38	0.75	0.55	1.30

Subject Number	Proficiency Level	Feedback Treatment	Topic Choice	Content Score	Organization Score	Vocabulary Score	Content/ Discourse Adequacy Score	Grammar Score	Mechanics Score	Grammar/ Mechanics Accuracy Score
76	"	Self	2	21	14	15	50	0.80	0.55	1.35
77	"	"	3	17	14	14	45	1.32	0.26	1.58
78	"	"	3	14	10	11	35	1.53	0.16	1.69
79	"	"	3	18	13	13	44	1.89	0.36	2.25
80	"	"	3	13	8	10	31	2.36	0.43	2.79
81	"	"	3	20	16	14	50	2.65	0.48	3.13

BIBLIOGRAPHY

Arapoff, Nancy. 1969. Discover and transform: a method of teaching writing to foreign students. TESOL Quarterly 3, 4:297-303.

Arnold, Louis V. 1963. Effects of frequency of writing and intensity of teacher evaluation upon performance in written compos tion of tenth-grade students. Unpublished Ph.D. dissertation. Tallahassee, Florida State University.

Beach, Richard. 1979. The effects of between-draft teacher evaluation versus student self-evaluation on high school students' revising of rough drafts. Research in the Teaching of English 13, 2: 111-119.

Braddock, R, R. Lloyd-Jones and L. Schoer (Eds.) 1963. Research in Written Composition. urbana: National Council of Teachers of English.

Brereton, John C. 1978. A plan for writing. New York: Holt, Rinehart and Winston.

Brière, Eugene J. 1966. Quantity before quality in second language composition. Language Learning 16, 3 & 4: 141-151.

Britton, J., T. Burgess, N. Martin, A. McLeod and H. Rosen. 1975. The development of writing abilities (11-18). London: MacMillan Education.

Buxton, Earl W. 1958. An experiment to test the effects of writing frequency and guided practice upon students' skills in written expression. Unpublished Ph.D. dissertation. Stanford, Stanford University.

Cardelle, Maria and Lyn Corno. 1981. Effects on second language learning of variations in written feedback on homework assignments. TESOL Quarterly 15, 3:251-261.

Calkins, Lucy M. 1978. Writers need readers, not robins. Language Arts 55, 6:704-707.

Celce-Murcia, Marianne. 1974. Report of an informal classroom experiment on speedwriting with a suggestion for further research* In T.P. Gorman (Ed.) Workpapers in teaching English as a second language, Vol. VIII. Los Angeles: University of California, Los Angeles.

Chaudron, Craig. 1977. Descriptive model of discourse in

the treatment of learners' errors. Language Learning 27, 1:29-46.

Chaudron, Craig. 1984. Evaluating writing: effects of feedback on revision. RELC Journal 15, No.2.

Clifford, John. 1981. Composing in stages: The effects of a collaborative pedagogy. Research in the Teaching of English 15, 1:37-53.

Clouse, Barbara F. 1983. Writing. New York: McGraw-Hill.

Cooper, C.R. 1977. Holistic evaluation of writing. In C.R. Cooper and L. Odell (Eds.) Evaluating writing: describing, measuring, judging. Urbana: National Council of Teachers of English.

Diederich, P. 1974. Measuring growth in writing. Urbana: National Council of Teachers of English.

Edelsky, Carole. 1982. Writing in a bilingual program: the relation of L1 and L2 texts. TESOL Quarterly 16, 2:211-228.

Elbow, Peter. 1973. Writing without teachers. New York: Oxford University Press.

Ellman, Neil, 1975. Peer evaluation and peer grading. English Journal 64, 3:79-80.

Emig, Janet. 1971. On teaching composition: some hypotheses as definitions. Research in the Teaching of English 1, 2: 127-135.

Emig, Janet. 1971. The composing process of twelfth graders. Urbana: National Council of Teachers of English.

Endicott, A.L. 1973. A proposed scale of syntactic density. Research in the Teaching of English 7, 1:5-12.

Erazmus, Edward T. 1960. Second language composition teaching at the intermediate level. Language Learning 10, 1 & 2:25-31.

Evolva, Jill, Ellen Mamer and Becky Lentz. 1980. Discrete point versus global scoring for cohesive devices. In John W. Oller Jr. and Kyle Perkins (Eds.) Research in language testina. Rowley: Newbury House.

Farrel, Kevin J. 1977. A comparison of three instructional approaches for teaching written composition to high school juniors: teacher lecture, peer evaluation, and group tutoring. Unpublished Ed. D. dissertation. Boston: Boston University-

Fellows, J.E. 1936. The influence of theme reading and theme correction on eliminating technical errors in written composition of ninth grade pupils. Unpublished Ph.D. dissertation. Iowa City, University of Iowa.

Ferguson, George A 1981. Statistical analysis in psychology and education. 5th ed. New York: McGraw-Hill.

Flahive, Douglas E. and Becky G. Snow. 1980. Measures of syntactic complexity in evaluating ESL compositions. In John W. Oller Jr. and Kyle Perkins (Eds.) Research in language testing. Rowley: Newbury House.

Flower, Linda S. 1979. Writer-based prose: a cognitive basis for problems in writing. College English 41, 1:19-37.

Flower, Linda and John Hayes. 1979. Identifying the organization of the writing process. In Lee W. Greaa and Erwin R. Steinberg (Eds.) Cognitive processes in writing. Hillsdale: Lawrence Erlbraum Associates.

Ford, Bob W. 1973. The effects of peer editing/grading on the grammar-usage and theme-composition ability of college freshmen. Unpublished Ed. D. Dissertation. Norman, University of Oklahoma.

Gaies, Stephen J. 1980. T-unit analysis in second language research: application, problems and limitations. TESOL Quarterly 14, 1:53-60.

George, H.V. 1972. Common errors in language learning: insights from English. Rowley: Newbury House.

Gorman, Thomas P. 1979. The teaching of composition. In Marianne Celce-Murcia and Lois McIntosh (Eds.) Teaching English as a second language. Rowley: Newbury House.

Hairston, Maxine. 1982. The winds of change: Thomas Kuhn and the revolution in the teaching of writing. College Composition and Communication 33, 1:76-88.

Hansen, Barbara. 1978. Rewriting is a waste of time. College English 39, 8:956-960.

Haswell, Richard H. 1983. Minimal marking. College English 45, 6:600-604.

Hatch, Evelyn and Hossein Farhady. 1982. Research design and

statistics for applied linguistics. Rowley: Newbury House.

Homburg, Taco J. 1984. Holistic Evaluation of ESL compositions: can it be validated objectively? TESOL Quarterly 18, 1:87-108.

Hughey, James B, Deanna R. Wormuth, V. Faye Hartfiel and Holly L. Jacobs. 1983. Teaching ESL composition: principles and techniques. Rowley: Newbury House.

Hunt, K.W. 1965. Grammatical structures written at three grade levels. NCTE Research Report No.3. Urbana: National Council of Teachers of English.

Hunt, K.W. 1977. Early blooming and late blooming syntactic structures, In C.R. Cooper and L. Odell (Eds.) Evaluating writing: describing. Urbana: National Council of Teachers of English.

Jacobs, Holly L., Stephen A. Zingraf, Deanna R. Wormuth, V. Faye Hartfiel and Jane B. Eughey. 1981. Testing a practical approach. Rowley: Newbury House.

Jacobs, Suzanne E. 1982. Composing and coherence. Linguistics and Literacy Series 3. Washington D.C.: Center for Applied Linguistics.

Kaczmarek, Celeste M. 1980. Scoring and rating essay tasks. In John W. Oller Jr. and Kyle Perkins (Eds.) Research in language testing. Rowley: Newbury House.

Karengianes, Myra L., Ernest T. Pascarella and Susanna W. Pflaum. 1980. The effects of peer-editing on the writing proficiency of low-achieving tenth grade students. The Journal of Educational Reseach 73, 4:203-207.

King, Jeane A. 1979. Teachers' comments on students' w..ing: a conceptual analysis and empirical study. Unpublished Ph.D. dissertation. Ithaca, Cornell University.

Knapp, Donald. 1972. A focused, efficient method to relate composition correction to teaching aims. In H. Allen and R. Campbell (Eds.) Teaching English as a second language. New York: McGraw-Hill.

Knoblauch, C.H. and Lil Brannon. 1981. Teacher commentary on student writing: the state of the art. Freshman English News 18, 2:1-4.

Lagana, Jean R. 1974. The development, implementation,

and evaluation of a model for teaching composition which utilizes individualized learning and peer grouping. Unpublished Ph.D. dissertation. Pittsburgh, University of Pittsburgh.

Lamberg, Walter. 1980. Self-provided and peer-provided feedback. College composition and communication XXXI, 1:63-69.

Larsen-Freeman, Diane and Virginia Strom. 1977. The construction of a second language acquistion index of development. Language Learning 27, 1:123-134.

Lim, Ho-Peng. 1983. Using T-unit measures to assess writing proficiency of university ESL learners. RELC Journal 14, 2:35-43.

Linn, William L. 1976. Contrastive approaches: An experiment in pedagogical technique. College English 38, 2:144-152.

Llyod-Jones, R 1977. Primary trait scoring. In C.R. Cooper and L Odell (Eds.) Evaluating writing: describing, measuring, judging. Urbana: National Council of Teachers of English.

Lyons, Bill. 1981. The POP method of responding to writing. English Journal 70, 3:42-43.

Maize, Ray C. 1954. Two methods of teaching English to retarded college freshmen. The Journal of Educational Psychology 45, 1:22-28.

Marzano, R.J. and S. Arthur. 1977. Teacher Comments on student essays: it doesn't matter what you say. ERIC Document ED 147 864.

Mattson, Marylu, Sophia Leshing and Elaine Levi. 1979. Help yourself: a guide to writing and rewriting. 2nd ed. Columbus: Charles E Merrill.

McColly, William. 1963. Comparative effectiveness of composition skills learning activities in the secondary school. Research Project No.1528. Madison: University of Wisconsin.

Moffett, James. 1968. Teaching the universe of discourse. Boston: Houghton Mifflin.

Murray, Donald. 1968. A writer teaches writing: a practical method of teaching composition. Boston: Houghton Mifflin.

Murray, Donald. 1972. Teaching writing as process, not product. In Richard L. Graves (Ed.) Rhetoric and

composition. Rochelle Park: Hayden.

Nold, E.W. and S.W. Freedman. 1977. An analysis of readers' responses to essays. Research in the Teaching of English 11, 2:164-174.

O'Donnell, Roy C., W.J. Griffin and R.C. Norris. 1967. Syntax of kindergarten and elementary school children: a transformational analysis. NCTE Research Report No.8. Champaign-Urbana: National Council of Teachers of English.

Partridge, Kathryn L. 1981. A comparison of the effectiveness of peer vs. teacher evaluation for helping students of English as a second language to improve the quality of their written composition. MA thesis. Honolulu: English as a Second Language Department, University of Hawaii.

Pellegrino, Victor C. 1982. You can write: practical writing skills for Hawaii. Honolulu: The Bess Press.

Perl, Sandora. 1978. Five writers writing: case studies of the composing process of unskilled college writers. Unpublished Ed. D. dissertation. New York, New York University.

Pfeiffer, Jerilyn K. 1981. The effects of peer evaluation and personality on writing anxiety and writing performance in college freshmen. Unpublished Ed.D. dissertation. Austin, Texas Tech University.

Pierson, Howard. 1967. Peer and teacher correction: a comparison of the effects of two methods of teaching composition in grade nine English classes. Unpublished Ph.D. dissertation* New York, New York University.

Pierson, Howard. 1972. Teaching writing. Englewood cliffs, N.J.: Prentice Hall.

Putz, Joan M. 1970. When the teacher stops teaching -- an experiment with freshman English. College English XXXII, 1:50-57.

Raimes, Ann. 1979. Problems and teaching strategies in ESL composition. Language in Education: Theory and Practice No.14.

Rivers, Wilga M. and Mary S. Temperley. 1978. A practical guide to the teaching of English as a second or foreign language. New York: Oxford University Press.

Rivers, Wilga M. 1981. Teaching foreign-language skills. 2nd ed. Chicago: University of Chicago Press.

Roberts, Paul. 1956. Understanding English. New York: Harper and Brother.

Sager, Carol. 1973. Improving the quality of written composition through pupil use of rating scale. Paper represented at the Annual Meeting of the National Council of Teachers of English. ERIC Document ED 089 304.

Schoen, Carol, Elaine Avidon, Nila Gandhi and James Vaughn. 1982. The writing experience. 2nd ed. Boston: Little, Brown and Company.

Scott, M, and G.R. Tucker. 1974. Error analysis and English-language strategies of Arab students. Language Learning 24, 1:69-97.

Shaughessy, Mina. 1977. Errors and expectations. New York: Oxford University Press.

SPSS Inc. 1983. SPSSX User's Guide. New York: McGraw-Hill.

Sullivan, Kathleen E 1980. Paragraph practice. New York: MacMillan.

Sutton, Joseph T, and Eliot R Allen. 1964. The effect of practice and evaluation on improvement in written composition. Cooperative Research Project No.1993. Deland, Stetson University. ERIC Document No ED 001 274.

Tate, Gary, and Edward P.J. Corbett (Eds.) 1981. The writing teacher's sourcebook. New York: Oxford University Press.

Taylor? Barry P. 1981. Content and written form: a two-way street. TESOL Quarterly 15, 1:5-13.

Voss, Ralph F. 1983. Janet Emig's the composing process of twelfth graders: a reassessment. College Composition and Communication. XXXIV, 3:278-283.

Watson, Cynthia B. 1982. The use and abuse of models in the ESL writing class. TESOL Quarterly 15, 1:5-13.

Witbeck, Michael C. 1976. Peer correction procedures for intermediate and advanced ESL composition learners. TESOL Quarterly 10, 3:321-326.

Witte, Stephen P. and Anne S. Davis. 1982. The stability of T-unit length in the written discourse of college freshmen: a second study. Research in the Teaching of English 16, 1:71-84.

Wolter, Daniel W. 1975. Effects of feedback on performance

on a creative writing task. Unpublished Ph.D. dissertation. Ann Arbor, University of Michigan.

Young, Richard. 1978. Paradigm and problem: needed research in rhetorical invention. in C.R. Cooper and L. Odell (Eds.) Research in composing. Urbana: National Council of Teachers of English.

Zamel, Vivian. 1976. Teaching composition in the ESL classroom: what we can learn from research in the teaching of English. TESOL Quarterly 10, 1:67-76.

Zamel, Vivian. 1982. Writing: the process of discovering meaning. TESOL Quarterly 16, 2:195-209.

Zamel, Vivian. 1983. The composing processes of advanced ESL students: six case studies. TESOL Quarterly 17, 2:165-187.

Zhang, Shuqiang and Peter Halpern. 1984. The effects of corrective ——— m a ——— quality of ESL students' compositions. MS. Honolulu: English as a Second Language Department, University of Hawaii.

Ziv, Nina. 1981. The effect of teacher and peer comments on the writing of four college freshmen. Unpublished Ed.D dissertation. New York, New York University.