

The connection between peer and self-evaluation  
in English as a second language composition

by

Audrey Popson Debje

A Thesis Submitted to the  
Graduate Faculty in Partial Fulfillment of the  
Requirements for the Degree of  
MASTER OF ARTS

Department: English  
Major: English (Teaching English as a Second  
Language/Linguistics)

Signatures have been redacted for privacy

---

Iowa State University  
Ames, Iowa  
1994

For Michael George Debÿe, my husband, who gently, but firmly, prodded me to finish this thesis: *Ich liebe dich*.

And for Marcia Sola, faculty adviser of the Loras International Student Association, without whose inspiration I would not be in this field today.

## TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
Peer Review	1
Self-Monitoring	2
Accuracy and Specificity	2
Lack of Relevant Research	5
Questions for This Study	6
Organization of the Study	8
CHAPTER 2. LITERATURE REVIEW	9
Cooperative Learning	9
Process Approach to Writing	13
Peer Response Issues	19
L1 and L2 peer groups: Some important differences	20
Development of accuracy in peer evaluation	28
Importance of task-specific questions	30
Common procedures for peer evaluation	33
The Role of Self-Evaluation	35
Rationale for self-evaluation	36
Common procedures for self-evaluation	39
Student Perspectives of Evaluation	41
Chapter Summary	43
CHAPTER 3. METHOD AND MATERIALS	44
Rationale for the Research Design	44
Background and Objectives for English 101C	46
Subjects	48
Instrumentation	48
Procedure	52
Scoring of Peer and Self-Evaluation Tasks	55
Student examples	56
CHAPTER 4. RESULTS	58
Results of Peer and Self-Evaluation Tasks	60
Explanation of statistical analyses	60
Accuracy results	62
Specificity results	64
Comparison of accuracy and specificity results	66
Total peer and self-evaluation task results	69
Analysis of Evaluation Question 3	70
Accuracy results	71
Specificity results	72

Comparison of accuracy and specificity results	74
Follow-up Questionnaire Results	76
Student perceptions and actual performance	77
Sample student responses	80
Chapter Summary	81
CHAPTER 5. CONCLUSIONS	83
Overall Conclusions	83
Implications for Composition Teachers	87
Suggestions for Further Research	90
REFERENCES	92
ACKNOWLEDGMENTS	103
APPENDIX A. PEER/SELF-EVALUATION QUESTIONNAIRE	104
APPENDIX B. FOLLOW-UP QUESTIONNAIRE	105
APPENDIX C. ASSIGNMENT SHEET FOR SURVEY REPORT	106
APPENDIX D. SCORE SUMMARY FOR ALL SUBJECTS	107
APPENDIX E. SCATTERPLOTS OF VARIABLES	108

**RECEIVED**  
SEP 20 1994  
IOWA STATE UNIVERSITY  
OFFICE OF GRADUATE ENGLISH

## CHAPTER 1. INTRODUCTION

For the last two decades, the process approach to writing has been touted as a great boon to writing instructors and their students. This approach grew out of opposition to a product-centered approach in which teachers "stressed expository writing, made style the most important element in writing, and maintained that the writing process is linear, determined by writers before they start to write" (Connor, 1987 p. 677). Many definitions have been written to describe the process approach, but essentially

the process-centered paradigm. . . focuses on writing processes; teaches strategies for invention and discovery; considers audience, purpose, and context of writing; emphasizes recursiveness in the writing process; and distinguishes between aims and modes of discourse (e.g., expressive, expository, persuasive; and description, narration, evaluation, and classification). (ibid.)

This thesis looks at one piece of this process: evaluation. Specifically, peer and self-evaluation of student writing in an English as a second language (ESL) composition course are examined. Also investigated are the issues of response accuracy, response specificity, and student perceptions of their ability to conduct peer and self-evaluation.

### Peer Review

Part of the process approach involves the use of peer review groups to help student writers improve their writing skills, such as those mentioned above by Connor, throughout the process of writing an essay. That is, students read rough drafts of each other's papers and make suggestions for improvement in content, organization, mechanics, and style "so that the student writers may acquire a wider sense of audience and work toward improving their

compositions" (Nelson & Murphy, 1993 p. 135). Research in education concerning cooperative/collaborative learning (e.g., Johnson & Johnson, 1987), in addition to the shift from an emphasis on product to an emphasis on process in the teaching of composition (Hairston, 1982), has supported the rationale behind the peer review stage of the process approach. Therefore, this step in the writing process is one that has received special attention from many composition teachers, and, in fact, an entire recent issue of the Journal of Second Language Writing (September 1992) has been devoted to this very topic.

### **Self-Monitoring**

Another of the goals of the process approach is for students to build their self-monitoring skills; i.e., students should not only rely on their teachers or peers for suggestions, but also learn to rely on their own judgments about their own writing. The combined techniques of peer evaluation and self-evaluation are, thus, essential ingredients of the process approach to composition. Some researchers claim that writers do become better self-evaluators over time as a result of first becoming good peer evaluators. For example, composition textbooks and teachers now provide student writers with checklists to be completed as a self-guide during the drafting stages of the writing process.

### **Accuracy and Specificity**

A major concern of composition teachers is that their students learn to give accurate and specific suggestions to their peers and to themselves in addition to gaining confidence in the social skills required in small-group interactions. Theoretically, over the course of a semester, students will become more adept and accurate at identifying specific strengths and weaknesses in their

peers' and their own rough drafts; however, there seems to be little empirical evidence for this in the literature on L1 and L2 composition. Many articles discuss how to devise better ways for students to conduct the peer review process (giving oral feedback instead of written comments, for example); yet, while these suggestions have been helpful for teachers in structuring the procedure, it is uncertain whether students really do become more accurate and specific in giving feedback over time. The success of peer and self-evaluation depends in part on whether students are able to be accurate and specific in their suggestions; otherwise, the value of peer and self-evaluation diminishes.

Another issue that is important to composition teachers when structuring the peer and self-evaluation processes is deciding on the type of guide questions to provide their students. Some composition teachers suggest using the same generic questions for each assignment, while others advocate using task-specific questions to elicit specific responses. One could suppose that a general peer review guide question such as "What did you like best about this paper?" would lead to a general response, whereas a specific guide such as "Identify where the writer gives support for his or her thesis statement" would elicit a more specific response. It seems logical to conclude that students who are given task-specific guide questions will have an easier time finding specific points to evaluate. Again, the argument remains to be made that this is entirely true, and this thesis, in part, examines this issue.

Response accuracy and response specificity are important to explore because these are factors that can determine the success or failure of the peer and self-evaluation tasks. That is, if students do not develop their evaluation skills so that they are accurate and specific in their feedback, then perhaps how we

teach the peer and self-review processes needs to be revised. Furthermore, if student perceptions of their evaluation ability do not match their actual performance, they will provide not only inaccurate and non-specific feedback, but will also be unaware of what they are actually doing and how to improve their skills.

Traditionally, the function of peer review sessions is to help one's classmates develop their papers more thoroughly, but are there benefits for students when evaluating their own writing? For example, if a student can accurately find certain types of problems, such as content and organization flaws, in her peer's essays, shouldn't she also be able to accurately find specific problems in her own essays? Shouldn't self-evaluation function similarly to peer evaluation in that a student writer becomes an accurate evaluator for both a peer and herself? Many composition teachers, including the writer of this thesis, have mixed feelings about peer and self-evaluation because the processes do not always produce the desired results. Even with training and practice in using peer and self-evaluation, students do not always give adequate peer responses, nor do they seem to improve their drafts when they receive accurate suggestions from their peers. Moreover, students do not seem to be able to find the same types of problems in both a peer's essay and their own, even when asked to look for specific elements. The reasons for this remain to be discovered, but White and Arndt (1991), among others, claim that "by learning to evaluate others' writing and responding in turn to evaluation of their own, students will gradually build up that capacity for self-assessment which is such a vital element in the process of writing" (p. 117). However, many writers and composition teachers, including White and Arndt, admit it is difficult for writers to see their work objectively

when they have been immersed in it. Although this point is an important one to research, it is beyond the scope of this thesis and will be treated only peripherally.

### **Lack of Relevant Research**

Composition research and cooperative learning studies primarily involving native (or L1) speakers of English focus on the social benefits and functions of peer groups, but few discuss specific details about how peer review helps one become a stronger self-monitor during the writing process. That is, many of these studies focus on how peers help other peers improve, but not how an individual acting as a "self-peer" helps herself improve her own writing. Again, many composition textbooks (and teachers) do provide checklists of self-guiding questions for a writer to use before and during the writing a first draft (e.g., Leki, 1989; White & Arndt, 1991), but there seems to be little written about the effectiveness of these self-monitoring devices in improving the writer's work after she has written a draft. First- language composition studies also provide a theoretical and anecdotal foundation for the importance of peer evaluation, as well as discussion of how to conduct peer review procedures so that they are more cooperative, but few show how peer evaluation directly helps a writer improve her self-assessment skills other than to say that a writer becomes more "aware" of her audience through reading her peers' comments. The purpose of including L1 peer evaluation research in this thesis is not to discredit it, but to show the background for second language peer evaluation research.

Some second language (L2) researchers who are concerned with peer evaluation have investigated other questions addressing the effectiveness of peer

review by using empirical studies, but L2 studies are also replete with anecdotal information and few address the issue of peer evaluation as a technique to help improve self-evaluation. Nelson and Murphy (1993), for example, are concerned with detailing the differences between L1 and L2 student writers and their use of peer evaluation. The researchers attempt to measure successful peer response group interactions in terms of tone (cooperative, competitive), accurate student marking of strengths and weaknesses on peers' papers, and consideration of peer comments in a student's own work (p. 140), but they do not include development of self-monitoring in their list of successes. They do address the fact that students often incorporate peer responses, some of which may be detrimental to their papers, but Nelson and Murphy do not address how individual writers can learn to judge the validity of their peers' comments for themselves. Furthermore, they do not discuss how having strong self-evaluation skills may aid students in making accurate judgments about their own work, a factor that is important under the process approach.

### Questions for This Study

To address some of the above issues, the research questions and null hypotheses that I set out to investigate and confirm are the following:

1. When conducting peer and self-evaluation of a draft, are student writers accurate in giving feedback for content and organization?

Hypothesis 1: Students are neither more nor less accurate in providing feedback to themselves than to their peers.

2. When asked to answer task-specific questions on peer and self-evaluations, do students provide specific feedback?

Hypothesis 2: Students are neither more nor less specific when responding to task-specific questions for themselves than when doing so for their peers .

3. Is there a connection between students' accuracy scores and their specificity scores for both peer and self-evaluation?

Hypothesis 3a: Students are neither more accurate nor more specific when conducting peer evaluation.

Hypothesis 3b: Students are neither more accurate nor more specific when conducting self-evaluation.

4. At which task are students better overall: peer evaluation or self-evaluation?

Hypothesis 4: Students are neither better nor worse when conducting peer evaluation overall than they are when conducting self-evaluation overall.

In addition to the questions above, a related question and research hypothesis investigates students' awareness of their ability to give feedback.

5. Is there a relationship between students' perceptions of their ability to conduct peer and self-evaluation and their actual performance on these tasks?

Hypothesis 5a: Students' perceptions of their ability to conduct peer evaluation are neither higher nor lower their actual performance on this task.

Hypothesis 5b: Students' perceptions of their ability to conduct self-evaluation are neither higher or lower their actual performance on this task.

This thesis attempts to answer these questions and confirm these hypotheses through a descriptive classroom study conducted in an Iowa State University English as a Second Language (ESL) composition course in which students completed both a peer evaluation and a self-evaluation of a rough draft of a writing assignment.

### Organization of the Study

Four other chapters are included in this thesis. Chapter 2 will describe how peer evaluation in both L1 and L2 composition came to be accepted and used as teaching and learning techniques. The chapter will also review articles explaining how teaching students to become better self-monitors during the writing process is just as crucial as teaching them to be good peer evaluators. In addition, this chapter will detail a variety of peer and self-evaluation procedures in both L1 and L2 composition classes and will focus on the importance of using teacher-provided task-specific guiding questions for these processes. Chapter 3 will discuss the methods and materials used for data collection in this particular thesis. Chapter 4 will discuss and analyze the results of this study, and Chapter 5 will present conclusions, implications for L2 composition teachers, and suggestions for further research.

## CHAPTER 2. LITERATURE REVIEW

This chapter begins with a review of the principles of cooperative learning and the process approach to teaching composition as background to the development of peer and self-evaluation as pedagogical techniques. The next section will cover relevant L1 and L2 peer evaluation research. Finally, the importance of self-evaluation in composition will be detailed.

### Cooperative Learning

Composition research regarding the value and usefulness of peer evaluation has drawn upon principles from education research regarding cooperative learning. One major source of information concerning the importance of cooperative learning comes from Johnson and Johnson's (1987) work Learning together and alone: Cooperative, competitive, and individualistic learning. In their text, they outline the validity of all these approaches with special emphasis on the value of cooperation. They note that cooperative learning is "as old as humankind" and that "the capacity to work cooperatively has been a major contributor to the survival of our species" (p. 10). In fact, Johnson and Johnson claim that the ability to cooperate is at the core of all successful human endeavors, including education. Johnson and Johnson's outline of the basic elements of cooperative learning spells out more thoroughly the requirements for successful cooperation:

1. Positive interdependence
  2. Face-to-face interaction among students
  3. Individual accountability for mastering assigned material
  4. Appropriate student use of interpersonal and small-group skills.
- (pp. 12-13)

These basic principles of cooperative learning can be applied to composition classes in the form of peer evaluation groups. For example, the element of positive interdependence is related to the notion that no one writes in a vacuum--that is, writers depend on their awareness of audience expectations in order to make their ideas clear. A writer is trying to communicate something that may be of interest or importance to a reader, and the reader responds to the writer to let her know if she has communicated her ideas successfully. The second element, face-to-face interaction among students, is crucial to the success of peer evaluation of rough drafts of compositions. While it is possible to anonymously critique another's work, Johnson and Johnson strongly stress that "it is the interaction patterns and verbal interchange among students promoted by the positive interdependence that affect educational outcomes" (p. 13). In terms of the peer review process, direct interaction among students has the potential to bring about desired results in one's writing because students are often impressed by their peers' opinions. Interdependence and interaction are desirable, but so is holding an individual accountable for achieving a level of success in her writing. Peer groups can give valuable feedback, but ultimately it is the writer who must make the final judgment about her own work. Student writers do not always have the opportunity to receive feedback on works in progress; thus, students must also monitor their own compositions in addition to monitoring that of their classmates. Finally, students must be trained in the "social skills needed for collaboration, and they must be motivated to use them" or else successful cooperation will not occur (p. 13). In the context of a writing class, teachers help students learn the social skills necessary for effective cooperation by setting up questions and guidelines for them to follow during

peer review sessions. Students learn that they must be active participants during peer evaluation in order for cooperation to take place.

Carson and Nelson (1994) also summarize points about the social benefits of collaborative learning from the works of Brown & Palinscar (1989), Bruffee (1986), Gere (1987), Kuhn (1970), Santos (1992), and Vygotsky (1962) all of whom believe that "knowledge is essentially a socially justified belief" and that it is "socially constructed" (Carson & Nelson, 1994 p. 18). These opinions follow Vygotsky's underlying conviction that a learner who interacts with another person, such as an instructor, in an exercise that would be too difficult for that learner to do on his own gains more knowledge or skill than an individual working alone. Carson and Nelson also cite Gere (1987 p. 75): "In the writing classroom, . . . cooperative learning is understood from a social perspective in which language is used not to convey ideas, but to develop them." Gere also states that "writing groups foster the development of writing abilities because they 'emphasize the communal aspects of intellectual life'" (p. 19). Writers have the further benefit of developing their sense of audience because they have real readers other than their teacher. The main point of all the articles cited by Carson and Nelson is that collaborative learning activities have a definite place in composition.

To further elaborate this point, it is important to cite Bruffee (1984), an L1 theoretical rhetorician, who builds a case for collaborative learning principles in composition classes. Bruffee believes that thought is an internal conversation one has with oneself, and writing is that internal conversation made public:

collaborative learning provides a social context in which students can experience and practice the kinds of conversation valued by college teachers. The kind of conversation peer tutors engage in with their tutees, for example, can be emotionally involved, intellectually and substantively

focused, and personally disinterested. There could be no better source than this of the sort of displaced conversation--writing--valued by college teachers. (p. 642)

Although collaboration is a necessary skill in the work place, it is one that "had been and largely still is ignored and hence wasted by traditional forms of education" (p. 638) In fact, Bruffee's philosophy is that "collaborative learning. . . harness[es] the powerful educative force of peer influence" yet remains underemployed by many educators. Bruffee also believes that learning and writing are social acts which need a social setting and that collaborative learning provides such a social context. That is, "a community of status equals--peers" (p. 642) is formed through collaborative learning. This is of particular relevance to writing classes, which also employ what Bruffee calls normal discourse. He defines this to be what a "community of knowledgeable peers . . . who accept, and whose work is guided by, the same paradigms and the same code of values and assumptions" (p. 642) engage in. Normal discourse involves everyone agreeing on the criteria for writing assignments--expected organization, relevant content. The end result of normal discourse is that all participants will have contributed to the discussion and that all are regarded with mutual respect. In other words, normal discourse is what is desired in a composition class. Addressing an audience of "status equals" is a primary goal in writing; thus, collaborative exercises like peer evaluation are key factors in achieving this goal.

It should be mentioned at this point that collaborative writing assignments and peer evaluation are processes that are related but not similar. That is, both employ the principles of cooperative learning outlined by Johnson and Johnson (1987), but the specific objectives for each are different. Collaborative writing assignments involve a group of two or more students

working on the same paper from conception to completion. Peer evaluation procedures are certainly a part of this technique, but it is possible to use peer evaluation for papers written by individuals, not groups, which is primarily how peer evaluation functions in many L1 and L2 composition classes. While the topic of collaborative writing assignments is of interest to both L1 and L2 researchers, it is beyond the scope of this thesis and will not be discussed.

### Process Approach to Writing

Hairston's (1982) article, "The winds of change: Thomas Kuhn and the revolution in the teaching of writing," is one of the hallmarks in the rise of the process approach to writing. As mentioned in the introduction of this paper, Hairston observed and called for a paradigm shift in how composition teachers treat writing instruction. She outlined this shift by showing where composition theorists had been and where they were going. The current-traditional paradigm holds the following principles:

it stresses expository writing to the virtual exclusion of all other forms, . . . posits an unchanging reality which is independent of the writer and which all writers are expected to describe in the same way regardless of the rhetorical situation, . . . neglects invention almost entirely, and . . . makes style the most important element in writing. (p. 78)

Current-traditionalists also believe that "competent writers know what they are going to say before they begin to write; thus their most important task when they are preparing to write is finding a form into which to organize their content" (p. 78). Other principles include the idea that "composing is linear . . . it proceeds systematically from prewriting to writing to rewriting" and "teaching editing is teaching writing" (p. 78). This approach is still clung to even though many writing teachers profess their emphasis on "process, not product." Hairston feels

that this view "denies that writing requires intellectual activity and ignores the importance of writing as a basic method of learning" (p. 79). For example, the current-traditional paradigm doesn't allow for small group conferences or development of self-evaluation abilities. Hairston feels that this is too limiting an approach.

A change in the paradigm came out of two decades of dissatisfaction with the current-traditional approach--a "revolution" which was a long time in coming. (p. 81). Hairston cites Young (1978) who shows that writing by trial and error (a process) and having that work criticized by others is how "thousands of people have learned to write" (p. 82). Hairston also cites Shaughnessy's philosophy from Errors and expectations (1977): students need to be allowed to make mistakes and develop a paper over a period of time as a process. Basic writers, especially, need time and process--they simply can not do what the current-traditional paradigm expects; it sets students up to fail. Hairston states:

Shaughnessy's insight is utterly simple and vitally important: we cannot teach students to write by looking only at what they have written. We must understand how the product came into being, and why it assumed the form that it did. We have to try to understand what goes on during the internal act of writing and we have to intervene during the act of writing if we want to affect its outcome. (p. 84)

Thus, one of the basic tenets of the process approach claims that "writing is an act of discovery" (p. 85), and this discovery process is necessary for all writers--skilled and unskilled alike--to develop their papers. As Hairston also states, many student writers may not know for certain what they want to write about until "their ideas develop in the process of writing" (ibid.). Composition teachers must also accept that for most students "the writing process is not linear"; rather, "it is messy, recursive, convoluted, and uneven. Writers write,

plan, revise, anticipate, and review throughout the writing process, moving back and forth among the different operations involved in writing without any apparent plan" (ibid.).

Since the publication of Hairston's article, many other practitioners have added their voices to the discussion of process approach, in both L1 and L2 studies, including Susser (1994), who outlines "discussions of L1 and L2 writing instruction [which] have been lively and at times contentious" (p. 31), none more so than those involving the process approach. He believes that the controversy over this approach stems from confusion as to what "process approach" really is, and in attempt to clear up some of this confusion, Susser delineates a difference between the writing process and process writing: the first term is used "when process is a synonym for the act of writing" and the second "to refer to process-based writing pedagogies" (p. 34). The first use of the term really indicates a variety of processes that are part of composing a paper--prewriting, drafting, reviewing, revising--rather than one overall process. The second term, Susser argues, is problematic because many composition theorists and practitioners attach the label "process" to their theories of writing, but Susser claims that "process pedagogies [are] not bound to any particular writing theory." In fact, "process is not the name of a writing theory" but "a component of most twentieth century writing theories" (p. 33). Further debate of this position is interesting and important for any discussion of writing pedagogies and theories, but it is beyond the scope of this thesis. However, the first use of the term is quite relevant to the topic of this thesis.

Susser cites Kostelnick (1989, p. 271), who highlights two fundamental parts of the process approach for L1, and by extension, L2: awareness and

intervention. This approach strives to make students aware that the act of writing involves a number of processes and that these processes vary among types of writing. In addition, Susser claims that students learn that "writing is often a process of discovery in which ideas are generated and not just transcribed" (p. 35); they also learn that they have control over their work: "choosing vocabulary, considering audience, and judging format" of their message. This awareness, ideally, makes students more responsible for their own writing, and gives them more control over their work.

The second primary element of the process approach, intervention, involves not only the teacher, but also the students. This includes procedures such as brainstorming, freewriting (which is commented on by others), journal writing, small group activities, student/teacher conferences, peer critiquing, and eventually, revising and editing. Susser cites Applebee (1986, p. 95), who says that these procedures are "designed to help students think through and organize their ideas before writing and to rethink and revise their initial drafts" (p. 35-36). That is, by giving guidance, the interaction between student writers and their teachers and peers can help them figure out what they mean to convey in their writing before a final, polished draft is due. As Susser succinctly explains, "intervention is not just the teacher's prerogative. Peer review and related procedures encourage intervention by classmates; the goal is for students to internalize this intervention as they write and revise" (p. 36). The claim is that by participating in activities such as those listed above, student writers will develop their own self-intervention skills. Whether or not this happens depends upon a variety of factors, some of which will be examined in Chapter 4 of this thesis.

Susser also details how the concept of process approach eventually appeared in L2 writing studies. Specifically, he notes how process has come to be employed by L2 composition teachers. Early L2 composition instruction was based on the audiolingual method and current-traditional rhetoric; thus, "writing was seen as grammar instruction, with emphasis on controlled composition, correction of the product, and correct form over expression of ideas" (p. 36). Students focused on grammar correction and patterning their writing after models instead of developing their own ideas and form. The "revolution" that Hairston described for L1 composition came into its own for L2 during the mid-1970s and early 1980s, and has had great effect on L2 composition research and pedagogy.

Zamel is given credit for indirectly introducing the process approach to L2 composition in 1976. She criticized the ESL field for focusing only on error correction, pattern writing, etc., instead of the expressive and creative processes of writing. (Susser p. 37). Others followed her lead in arguing against the use of models to teach writing, and the giving of severe penalties for errors, and they instead began following the principles of the process approach as they were defined in L1 research. Along with Zamel, Susser cites Kroll (1978) and Raimes (1979) with bringing process ideals into ESL composition research. From those early discussions of the process approach in L2 writing, many others have followed by including these principles in their writing pedagogies, such as McKay (1982) who focuses on prewriting strategies, Spack and Sadow (1982) who use journal writing, Spack (1984) who advocates teaching intervention techniques, and Krashen (1984) who encourages giving feedback during the drafting stages of a paper rather than at the end of the writing process (Susser, 1994 p. 37).

However, as Susser also mentions, proliferation of the discussion of process in L2 has met with resistance by those committed to audiolingualism or current-traditional practices. For example, Susser cites Barnes' (1983) criticism of what he believed was "ignoring of the product" and leading to personal writing and narratives, rather than academic papers. Zamel (1983b) counters by pointing out that her students used the process approach to create ideas for and to write research papers and academic essays and that process is valuable for all kinds of writing. Zamel (1982, 1983a, 1984b) has also been criticized by several other sources for her emphasis on process over product, including Reid (1984), who believes Zamel overlooks the end product of academic writing (e.g., research papers) in favor of the processes and stages of creative writing (e.g. personal narrative, poetry). Zamel (1984a) replies that "a process approach is by its very nature concerned with product" (Susser, 1994 p. 38) and that writing by nature is an act of creation. Hughey, Wormuth, Hartfiel, and Jacobs (1983) also defend this position: "all writing must be taught as creative because of the creative processes that make any and all writing possible" (p. 39). Because of these stances, process writing practices have "reached the mainstream of ESL writing instruction" (p. 39).

Diaz (1986b) also looks specifically at the process approach in ESL writing and generates some hypotheses about its effectiveness for ESL. She has found through her ethnographic study that the process approach gets away from rigid error identification and correction methods of teaching ESL composition and focuses on helping students develop overall strategies for writing, using such activities and techniques as free writing, peer writing groups, and peer group discussions on grammar and syntax, teacher conferences, student choice of topics,

rehearsal and invention strategies, emphasis on purpose and audience, and daily process journals. Diaz found that "not only are process strategies and techniques strongly indicated and recommended for ESL students, but also when used in secure, student-centered contexts, the benefits to these students can go far beyond their development as writers" (p. 14). The process oriented classroom enhances ESL students' audience awareness and self-esteem as writers, as well as their awareness of the writing process and their willingness to revise.

The debate surrounding the use of the process approach in L1 and L2 composition is far from over. However, the use of peer groups and other pedagogical techniques that follow from the process approach hold an important place in L1 and L2 composition.

### Peer Response Issues

The wealth of literature on the topics of peer and self-evaluation in L2 composition is vast and growing, and, as Rothschild and Klingenberg (1990) point out, these concepts are well known in L1 research and widely used in all levels of writing. The majority of research concerning peer and self-evaluation is theoretical and based on individual teacher experience, and in particular, focuses on how to make peer evaluation less threatening to students or how to make peer evaluation a more integral part of the writing process. The use of self-evaluation has also been touted as a way to make students more responsible for their writing. What is missing from the literature in both L1 and L2 is a sense of "how much" students improve in their evaluation abilities over time. At issue in this study, however, is not whether peer and self-evaluation are valid techniques; rather, an attempt is made to examine particular facets of these processes. Therefore, in order to tighten the focus of this section, only articles

that discuss the issue of developing accuracy over time, the use of task-specific guide questions, and various peer and self-evaluation procedures will be included. Important differences between evaluation procedures for L1 and L2 are highlighted first.

### **L1 and L2 peer groups: Some important differences**

As mentioned above, peer evaluation is a technique that has been increasingly included in L1 composition courses as a result of research in cooperative learning and the development of the process approach in writing. It should be mentioned, however, that the use of response groups is not a recently discovered technique. Rather, as DiPardo and Freedman (1988, 1987) explain, response groups have existed in American composition classrooms since the colonial period, and also received great attention during the 1900s and 1910s (p. 123). DiPardo and Freedman have also discovered in their extensive reviews of L1 research about peer response groups (their studies include five pages of references) that theoreticians and practitioners alike have been concerned primarily with the social benefits of peer evaluation, most specifically the nature of peer interactions and the amount of control students have over the peer review process. There are varied findings on these issues. For example, DiPardo and Freedman cite Gere and Abbott (1985), Gere and Stevens (1985), and Nystrand (1986) whose work shows that peer evaluation is overwhelmingly worthwhile, while Berkenkotter (1984) and Newkirk (1984a, 1984b) find several problems with peer review. DiPardo and Freedman summarize evidence which shows peer evaluation to be an effective tool in L1 composition. For example, they find that "response groups receiving fairly minimal guidance [from the teacher] are capable of staying on task" (p. 137) and that students' comments are

often as specific as the teacher's. Also, students who work with peer groups view revising and editing as separate processes. In fact, revising becomes "reconceptualization" (p. 138). Another positive finding is that the students learn to give each other accurate and specific feedback because they are engaged in the peer group and want to give and receive help from each other.

Other studies, such as Newkirk (1984a, 1984b), show that sometimes students are too helpful and that "strong peer identification among the students makes them more willing than their readers to fill in missing elaboration as they read, thus rendering them more tolerant of what the teachers consider thin or undeveloped prose" (cited in DiPardo and Freedman, p. 140). Also, students are too likely to impose their personal opinions on their peers' writing, thus misreading or rejecting ideas rather than helping the writer improve. Berkenkotter (1984) examines how students handle "the sometimes confusing task student writers face in reconciling their own imperatives with the suggestions of others" (cited in DiPardo and Freedman, p. 140). Through case studies, Berkenkotter identifies the distinct reactions to peer evaluation of three students: one who absolutely refused to consider peer suggestions, one who kept her focus and purpose in spite of unclear feedback, and one who was over-influenced by "the sometimes hypercritical feedback" (p.141) given and who lost sight of her purpose. The range of reaction among these students leads Berkenkotter to caution that peer response must be set up very carefully because "the interplay of 'subtle emotional and intellectual factors' can leave some students feeling more confused than enlightened" (p. 141). That is, students' sense of authority over their own work may not be clearly defined in peer response sessions.

Other L1 practitioners argue over the issues of student control and the structure of the peer response task. DiPardo and Freedman cite Hillocks (1984), in addition to Newkirk and Berkenkotter, who all argue the need for a structured approach to peer evaluation (giving students specific tasks, for example) and Nystrand who allows his students to follow an Elbowian approach, responding viscerally and idiosyncratically. Overall, nevertheless, DiPardo and Freedman argue for the continuation of peer response groups in L1 because they foster a wider sense of audience for student writers, and they aid in building trust among classmates.

Although L1 studies argue for the use of peer evaluation groups, Nelson and Murphy (1993) point out that these findings "do not necessarily apply to L2 students" (p. 135). They find that L2 students differ in two important ways: "Because L2 students are in the process of learning English, they may mistrust other learners' responses to their writing and, therefore, may not incorporate peer suggestions while revising." Also, "L2 students who view the teacher as 'the one who knows' may ignore the responses of other students, not merely because English is the respondents' L2, but because of the perception that fellow students are not knowledgeable enough to make worthwhile comments about their work" (p. 136).

These concerns underscore the need of L2 composition teachers to clearly explain the purpose and goals of peer evaluation to students who might otherwise think it a pointless exercise. Also, students need to realize that they and their classmates have the ability to make accurate and specific suggestions for improvement, but as Leki (1990) cautions, "how can an inexperienced ESL writer know what to accept and what to reject from among the comments made by

another inexperienced ESL writer/reader?" (p. 11). The answer seems rather simple: students must have training and several opportunities to practice peer evaluation. Stanley (1992) cites Allaei and Connor, 1990; Flynn, 1982; and Leki, 1990, all of whom "urge teachers to consider the ways in which they prepare students to become peer evaluators" (p. 219). Without adequate preparation and practice in giving accurate and specific feedback, students will be unsuccessful in performing this task. Stanley found in her study that when students receive coaching and training for peer review, they "demonstrate a greater level of student engagement in the task of evaluation, more productive communication about writing, and clearer guidelines for the revision of drafts" (p. 219). In fact, the students who received coaching were made aware that they could trust their peers' comments and that they should seek out help from their classmates.

While the issue of training students to be effective peer evaluators is one that pertains to both L1 and L2 students, there are cross-cultural differences that must also be taken into account if L2 writers are to see peer evaluation as valuable. For example, Carson and Nelson (1994) examine some of the cross-cultural differences among ESL writers in their perceptions of the peer review process and how these differences affect the process. Although their primary purpose is to examine the dynamics of peer group relations, they do address how different cultures view collaborative work. Students coming from "collectivist" (p. 17) cultures, such as Japan and China, may not see how peer response groups are meant to help individual writers as opposed to helping the whole. These students also are concerned with maintaining the harmony of the group (that is, not offending anyone or arguing), often putting aside the task of providing accurate and specific feedback on their peers' drafts, especially when the feedback

may be critical of their partners. Carson and Nelson also mention the categories of "ingroup" and "outgroup" (p. 24) as factors that impede the peer review process. If students are not totally committed to their "ingroup" or are ostracized for being in an "outgroup," the cohesion required for effective communication does not develop. Often "outgroup" is defined culturally or socially; thus, even if students belong to the same peer "ingroup" for the entire semester, there may be "outgroup" perceptions blocking the path to successful interaction.

Leki (1990), while not directly comparing L1 and L2 studies, notes that many L2 students differ from L1 students in their approach to peer evaluation because of the methods of writing instruction they have previously encountered. Many L2 writers have experienced composition classes focusing on "practical applications of grammar lessons" (p. 9) rather than on the writing processes. Thus, L2 students tend to confuse editing with responding. They fail to respond as readers, and instead "are likely to respond to surface concerns of grammar, mechanics, spelling, and vocabulary, taking refuge in the security of details of presentation rather than grappling with more difficult questions of meaning" (p. 9). Leki also suggests that students rely on editing because they are imitating responses they themselves have received from their teachers, who are, of course, considered by the students to be the only authorities of knowledge in the classroom.

Leki also echoes Nelson and Murphy's concern that students don't view their peers' comments as valid when, in fact, they are, but Leki also notes that sometimes these comments are not to be trusted because students do not always respond in valid and appropriate ways. As is true in L1 composition, L2 students

attempt to imitate teacher responses, but often give "rubber-stamp" comments that are also wrong:

pushed by this peer responding activity to say something, and at a loss as to what to say, students often resort to exhortations like 'Be specific' or 'Give some examples.' It is not unusual to see labels like this stamped in the margin of a paper at a spot where it is difficult to imagine what an example might be or how an example might clarify an assertion. (p. 10)

Leki notes that these types of inappropriate, inaccurate comments show how students may misunderstand the meaning of "Give examples" because in their own writing experience, they may have received the same inaccurate comment for similar problems in their own papers and have not learned how to judge this for themselves.

Whatever the differences between L1 and L2 composition, and the problems encountered by ESL writers, peer evaluation has earned a place in second language writing classes. For example, Mittan (1989) discusses the need to empower students by "highlighting and nurturing the strengths" they have in their writing (p. 207). To do this, he suggests that students need writing assignments that give them chances to draw on their own knowledge as they write and to show what they are able to do. Part of this idea includes having students focus on the strengths in their writing as a way to discover the weaknesses that exist. He contrasts this with what typically occurs in writing classes: students want teachers to correct their mistakes, and teachers oblige them. However, Mittan claims that shifting "the focus of our comments on student papers to what is working" will ultimately help students "revise what isn't successful" (p. 207). Doing so allows students to "harness their communicative power," which is one of the main goals of composition (p. 207).

To meet the goal of empowerment, Mittan explains why peer evaluation is vital to the development of an ESL student's "communicative power" and gives a "core list of benefits" of peer review for students and teachers (p. 209). By drawing on the research of Bruffee (1984), Halliday & Hassan (1985), and Moffett (1968), Mittan also argues that "language, language use, language learning, and learning in general" are social in nature and that language learning requires feedback from others because we direct our language to other people. More directly stated, we use language to communicate with others (p. 209). Using peer response allows students to receive feedback that may be more meaningful to them than teacher comments, and consequently this feedback helps students communicate their ideas more successfully. Mittan also gives five main reasons why peer reviews are so beneficial in L2 composition:

1. Peer reviews provide student writers with reactions, questions, and responses from authentic readers who provide a stronger motivation for revision.
2. Students receive more feedback from multiple and mutually reinforcing perspectives.
3. Because they reciprocate in the role of audience for their peers, students gain a clearer understanding of meeting the reader's needs. At the same time, by responding critically to their colleagues' writing, students exercise the critical thinking they must apply to their own work.
4. There is an affective element to peer interaction: students see that their peers also have difficulty writing and may gain confidence in, or at least feel less apprehension about, their own abilities.
5. Other practitioners have claimed that one advantage for the teacher is time saved because of a decreased reading load (pp. 209-210).

Of these, the first four points are the most directly related to the notion that cooperative learning opportunities are essential in empowering students to

become better writers, and are elements that are important for all writing teachers to consider. And of course, the fifth point is an added incentive for using peer evaluation.

Bell (1991) also provides compelling reasons why peer response groups should be used in ESL writing. He states that "writing groups are particularly suited to ESL" (p. 66). because they make use of all language skills: reading, writing, speaking, and listening. Writing groups are also "especially suited to certain orientations in the ESL field--for instance, communicative competence" (p. 66). Peer evaluation primarily gives students the opportunity to employ their language skills and to negotiate meaning when breakdowns in understanding occur.

Witbeck (1976) claims that using "peer-correction procedures results in increasingly more accurate and responsible written work on the part of most students and fosters a more constructive classroom atmosphere for teaching the correctional aspects of composition" (p. 321). Although Witbeck focuses mostly on errors in mechanics, grammar, spelling, he makes the case that peer evaluation is essential in L2 because there is a fundamental difference between the actions of proofreading and correcting errors and the process of learning the conventions of writing in a second language--teachers cannot assume that ESL students can do the former. Witbeck states that he has "always been skeptical as to how much good it does a student to see an error marked and then, alone at his desk, fix it up as best he can" (p. 321). In fact, "the correction of an error in a particular context does not . . . often lead to the elimination of the same kind of error in subsequent contexts, at least not without some kind of two-way discussion of the principle involved" (p. 321). Thus, even though Witbeck

addresses primarily issues of grammar and mechanics, his argument for the validity of peer evaluation as a means to achieving greater accuracy is important to note.

### **Development of accuracy in peer evaluation**

Whatever the procedure that a class follows in giving peer feedback, one important goal is that students will provide each other with accurate responses. Several L1 and L2 composition teachers and researchers claim that, over time, student accuracy in giving responses will improve, and it is important to investigate whether this is true. While the list of researchers that discuss this issue is quite extensive, (e.g., Carson & Nelson, 1994; Chaudron, 1983; Diaz, 1986a, 1986b; Hafernik, 1984; Hvitfeldt, 1986; Lamberg, 1980; Leki, 1990; Mittan, 1989; Nelson and Murphy, 1992/1993; Rothschild and Klingenberg, 1990; Stanley, 1992; White and Arndt, 1991; Witbeck, 1976, to name a few), this section focuses on the work of Jacobs and Zhang (1989) whose study investigates the extent to which accuracy in peer response improves and the factors which contribute to the development of accuracy. They raise a very important question regarding the use of peer feedback in L2 classrooms: "Do L2 learners provide mostly faulty feedback to their peers, miscorrecting rather than correcting composition drafts? Obviously, if the peer readers are unable to provide helpful suggestions for improvement, what is the point of using peer feedback at the revision stage?" (p. 2) In spite of all the attention given to the social benefits of peer evaluation, Jacobs and Zhang believe that there is not enough research about how accurate students are in giving feedback to their peers. Their study investigates a group of students in Thailand who were third-year English majors and who had previously conducted peer evaluation. The students were asked to read their

peers' papers and offer suggestions about improving the grammar and mechanics. Students made corrections by drawing lines through structures they believed to be incorrect and by writing what they felt was correct above them. Students also marked what they believed were errors without making corrections if they were uncertain about their judgments. Students conducted the peer feedback procedure on two drafts (first and second) of the same paper. This was done to determine if the students were revising according to their peers' suggestions. Jacobs and Zhang coded these corrections and markings into four categories:

A1. Original wrong--correction wrong

Example: (original) Suwit live in Chiang Mai.  
(correction) Suwit living in Chiang Mai.

A2. Original wrong--correction right

Example: (original) Suwit live in Chiang Mai.  
(correction) Suwit lives in Chiang Mai.

A3. Original correct--correction also right

Example: (original) Suwit lives in Chiang Mai.  
(correction) Suwit is living in Chiang Mai.

A4. Original correct--correction wrong

Example: (original) Suwit lives in Chiang Mai.  
(correction) Suwit live in Chiang Mai. (p. 4).

What Jacobs and Zhang discovered was that "by far the largest type of peer readers' corrections was A2, i.e., accurate correction of incorrect forms (74.68% for the first copy, 71.60% for the second copy)" (p. 5). They also looked at the frequency of students making unacceptable corrections (A1 and A4 combined): 16.45% and 18.52% for copy 1 and copy 2, respectively (p. 5). Finally, they also discovered that the frequencies of A3 and A2 combined "constitute 83.54% and 81.48%" for the two drafts (p. 5).

The results of this study indicate that "peer readers have quite an accurate sense of how sure they ought to feel about suggestions concerning grammar" and when "they are not sure, they are really not sure, and . . . they refrain from suggesting corrections" (pp. 5-6). Jacobs and Zhang concede that the proficiency level of their subjects may be a factor in their ability to provide mostly accurate feedback; however, they do not address whether students with lower writing proficiency are less able to provide accurate grammatical feedback. Thus, it is not clear whether Jacobs and Zhang's findings are pertinent to a larger population of L2 writers. It is also uncertain whether students are just as accurate in providing feedback for content and organization. Nevertheless, the main conclusion of the Jacobs and Zhang study, that "peer feedback does not seem to provide as much misleading guidance as some instructors and students fear" (p. 17), gives writing teachers an incentive to use peer evaluation with their students and to help their students become more accurate by providing them opportunities to practice their evaluation skills.

### **Importance of task-specific questions**

While DiPardo and Freedman (1988), along with several other L1 composition practitioners, argue for the idea of a "teacherless" peer review approach based on the ideas of Elbow (1973) and Macrorie (1979), this idea is not espoused by L2 composition instructors. A "teacherless" peer evaluation session "offer[s] only a bare minimum of guidance to groups, leaving students to devise strategies for responding to one another's writing that are largely intuitive and highly individual" (p. 127). As Nelson and Murphy (1993), among others have discovered, however, is that ESL students are often not equipped to be "teacherless" because they will flounder and not know how to respond

appropriately. Teachers want students to be able to give specific comments and suggestions for improvement to their peers, rather than bland generalizations. Several L2 composition teachers (e.g., Chaudron, 1983; Hafernik, 1984; Hvitfeldt, 1986; Jacobs, 1987; Mangelsdorf, 1992; Nelson and Murphy, 1992, 1992/1993; Stanley, 1992; Witbeck, 1976) provide students with task-specific questions instead of more general guiding questions in order to facilitate the peer review process, which to many ESL writers is a foreign and novel concept. This section focuses specifically on the work of Hvitfeldt (1986), Mittan (1989), Lamberg (1980), and Leki (1989), who advocate giving students specific evaluation tasks.

Hvitfeldt's (1986) experience with L2 peer evaluation finds that "if students are given very specific guidelines to follow, they often do a very credible job of analyzing the strengths and weaknesses of their peers' writing" (p. 1). She also finds that, even though L2 students "are not the best judges of grammaticality, word choice, and mechanics, they can develop critical abilities concerning the content and organization of an essay" (p. 2-3). Therefore, Hvitfeldt limits peer evaluation to these areas. Although this article presents a rationale for using specific peer critique guidelines based on her own experience teaching intermediate and advanced-level ESL composition, she fails to provide any empirical support for her claims. She does, however, agree with Mittan (1989) and Leki (1989) in their belief that peer review guidelines should and can be designed for each assignment, and in fact, Hvitfeldt provides examples to demonstrate how peer and self-review sheets can be developed.

Mittan (1989) also designs peer review sheets specifically for each paper assignment. He mentions the laments of his colleagues who use the same peer evaluation sheet for every paper and find that their peer review sessions

"flopped" (p. 215). To avoid this problem, Mittan states that it is worth the extra time to design peer evaluation questionnaires that will elicit "more productive student responses" by having students answer questions that are tailored to the current assignment. Instead of asking a question such as "What did you like the best about your partner's writing?", Mittan instructs students to specifically identify the most interesting idea and to give an explanation of why it was the most captivating. He does not allow students to respond with only "yes/no" answers--the more specific the suggestion, the better. Leki (1989) agrees with Mittan. In the instructor's manual to her text Academic writing: Techniques and tasks, she argues for the use of task-specific guide questions and provides a different set of questions for each writing assignment in her text.

Lamberg (1980) discusses the need for both qualitative and quantitative types of responses. For example, a qualitative response states: "this is an interesting comparison" or "this is an intriguing opening" (p. 64), and while this type of response can be specific, Lamberg believes that quantitative responses are more specific because they focus on particular paragraphs, sentences, and words. In Lamberg's opinion, quantitative comments help students develop more sophisticated papers because using these responses is a way for writers to measure and increase "syntactic variety." Students learn to identify "different types of sentence structures" that are used (p. 64). Individuals can gather quantitative information about their own papers by answering specific questions about their own papers using a checklist--a systematic tool.

This thesis attempts to make some definite conclusions about whether the use of task-specific questions really elicits specific answers. Most of the research cited here provides a theoretical basis for using specific peer critique questions or

explains teachers' personal experiences with them, but little has been supported statistically.

### **Common procedures for peer evaluation**

As mentioned above, the issues of accuracy and specificity are of key importance. In addition to these concerns, the methods that are used to conduct peer review sessions are also of great interest. For example, some teachers instruct students to sit together in small groups to discuss papers; others ask students to provide written feedback; still others advocate a combination of the two. Definitive research proving the effectiveness of one of these methods over the others has yet to be produced; in fact, teacher preference and experimentation often influence which method is chosen. However, discussion of the variety of approaches is considered here to give background for the approach taken by this study.

Mittan (1989) follows a procedure in which students spend one class period reading their peers' drafts and writing a response using a teacher-prepared, task-specific questionnaire which focuses on particular elements of the peers' papers. Mittan then has students exchange oral comments in order to elaborate on what they had written about their peers' work (p. 208). These peer response questionnaires are collected and returned with comments from the teacher. Also, after students have practiced peer response for several assignments, the instructor has them complete self-evaluation questionnaires, which are collected and commented on as well. In order to teach this process, Mittan strongly suggests that students be trained to become effective peer reviewers by first practicing in class with an anonymous essay. By doing this,

students "try out their review voices" and are guided by the instructor "toward useful responses" (p. 213).

Leki (1989a) follows a slightly different procedure. For each writing assignment in her text, she provides possible self-analysis questions and expects students to complete these for all writing assignments. These are normally to be answered after writing a rough draft and given in with the first draft of an essay, rather than done in class. Peer evaluation is completed during the class period. Leki provides guiding questions that are specific to each assignment. During the peer response session students sit together as they write their responses so they can talk to each other and ask clarification questions as they make their comments. After they finish writing, the students exchange responses, discuss points of agreement and disagreement, and then they take the comments home to analyze and incorporate changes in their drafts. Leki also takes in the peer and self-analysis responses and comments on them.

White and Arndt (1991) suggest that more open-ended methods of evaluation can be used for peer response. For example, their process involves peers reading each other's papers and recording their responses in their notes. Then, from these notes peers can construct letters to their partners that comment on their general likes and dislikes of the drafts (p. 124-128). Another procedure that White and Arndt follow is quite similar to Leki's, but the response questions that are used are not necessarily tailored to the particular assignment. Basically, students work with a partner in class. During that time, they read each other's drafts, making notes of places in the drafts that they specifically liked or enjoyed, specifically disliked or found unnecessary, found unclear, and would have liked to know more about (p. 130). Students also identify the main idea of the paper.

After they have finished writing comments, students then return the drafts to the writer to discuss and clarify their responses.

Thornburg and Schmidt (1994) use a round robin group technique in which groups of four to five students meet with the teacher to conduct the peer review. Each student reads her own paper aloud without interruptions from the classmates. After the student finishes reading, the members of the group take turns telling the writer what they would like to know more about in the paper, or what they think may be deleted, but there is no sheet of peer critique guidelines. Finally, the writer, who has been taking notes, has a chance to respond, either elaborating or defending her choices. There is no "cross-talk" allowed; that is, students are not allowed to interrupt or discuss until everyone has had a turn, then the writer can say if she wants to open the floor for further discussion.

These examples of peer evaluation procedures are by no means the only alternatives, but based on a survey of approaches, these are the most commonly used by both L1 and L2 classes.

### **The Role of Self-Evaluation**

Self-evaluation ideally "leads the student toward greater self-reliance and independence" (Beaven, 1977 p. 142). In much of the literature written about self-evaluation in composition, this type of statement appears, along with the hope that students will develop their self-evaluation skills over time as a result of becoming better peer evaluators. Students do have power and control over what they produce in writing; thus, they should be given the skills necessary to realize this fact for themselves. In fact, the topic of student writers practicing self-evaluation has been written about at length (e.g., Allen and Roswell, 1989;

Berkenkotter, 1983; Christensen, 1982; Cumming, 1986; Dicker, 1981; McKay, 1983; Otto, 1992; Rubin, 1985; Schwarte and Matsumura-Lothrop, 1987; Stern, 1992). By being trained to become strong self-evaluators, students ultimately will assume responsibility for assessing their writing; at least, this is what the literature suggests. However, White and Arndt (1991) claim that "the cultivation of a sense of responsibility for being one's own critic" (p. 116) is not stressed enough in writing classes, particularly in L2. Student writers learn to rely too heavily on the teacher and/or other students for assessment of their work. That revising is "what writing is all about" is a difficult concept for many students to understand because of this reliance on outside assessment, but White and Arndt stress that "we have to try to persuade our students that it is ultimately not the teacher but they themselves who must decide whether their text fulfills its intended goal. They have to be their own evaluators, for without a sense of what is wrong with a text, there is a little hope of being able to put it right" (p. 116).

White and Arndt also strongly believe that "students will gradually build up that capacity for self-assessment which is such a vital element in the process of writing" (p. 117) if they learn how to evaluate their peers' work and learn how to respond to peer feedback of their own papers. Whether students' self-evaluation skills develop as a result of learning how to be an effective peer evaluator, or, rather, because of direct training in self-evaluation is unclear in the literature. Thus, this thesis investigates the possible connection between the two.

### **Rationale for self-evaluation**

Beaven (1977) "gives a careful analysis of the rationale for and advantages/disadvantages of students' participation in the process of evaluation" (p. 134).

That is, "the individual student, not the researcher or the teacher" plays "a prominent role" in the evaluation of her own writing. Beaven begins by showing the types of questions students need to ask themselves as they write, questions similar to those students ask their classmates during the drafting stages (p. 143). Beaven wants students to learn that they and their teachers can find strengths and weaknesses in writing and that students can learn to be accurate and specific in how they evaluate themselves. "Writers must eventually become independent [and] able to identify passages that gracefully communicate meaning and intention; writers must become their own editors" (p. 144). Strong self-evaluators are, theoretically, able to distance themselves from their own writing and treat it as someone else's.

Although Beaven provides a rationale for student self-evaluation, Lamberg (1980) believes that many composition teachers (primarily before the publication of his article) have seen feedback as something that comes from a reader, the teacher or a peer, and not the writer herself. He notes that "research studies and discussions in articles and methods texts, typically, are limited to responses from the teacher and/or from peers" (p. 63). Lamberg also argues for the necessity of providing student writers opportunities to evaluate themselves. He acknowledges the importance of peer and teacher feedback, but argues that, concurrently, students should evaluate themselves.

Lamberg also claims that teachers do not always motivate students to improve their writing only by correcting errors; they should encourage students to attend to content and organization of developing papers as well. In addition, Lamberg argues that teachers often ask students to attend to too many errors at once, rather than focusing on areas which have been discussed in class. He gives

as an example of ineffective feedback: a teacher marks but does not correct spelling errors on a student's rough draft. In a few days, the student gives in a "revision" which "has as many errors" as the first draft because the student may not have learned why something was marked as an error or how to correct it. External teacher feedback has failed, but when "students are asked to proofread their papers for spelling and underline every word they have spelled correctly, then compute and record the total number of correctly spelled words" they "show an increase in the proportion of correctly spelled words while maintaining or increasing length" of the essay (pp. 63-64). Thus self-provided feedback has proved more effective in this case because the student had control over a specific task--correcting spelling errors--rather than being asked to correct any number of any type of error.

Lamberg has also found experimental research which "provides support for the idea that students can improve as a result of providing themselves and peers with feedback" (p. 64). The main conclusions of these studies are summarized as follows:

1. Two studies (Allen & Sutton, 1964; Pierson, 1972) "reported no statistically significant differences between groups' [sic] receiving peer and teacher-provided feedback, respectively" (p. 64).
2. Two studies (Lagana, 1974; Ford, 1973) "reported superior performance on compositions and grammar-usage test for groups receiving peer-provided feedback as compared to groups receiving teacher-provided feedback (though differences were not significant)" (p. 64).
3. Two studies (Millet, 1969; Sager, 1973) showing that "students who were taught to use a rating scale on their own and the [sic] peers' writing had significant differences over a no-feedback control group" (p. 64).

Lamberg also cites findings from Wolter (1975) that students who gave themselves feedback and students who received teacher response "both showed

significant gains in their writing from before and after the experiment" (p. 64). The results of the above studies indicate that it is better for students to receive some feedback on their work than none at all. Again, this is not at issue in this thesis; however, how accurately and specifically students give feedback is.

Lamberg states that students are absolutely able to provide themselves with feedback (p. 64). This is obviously an important objective of composition classes, since writers are ultimately responsible for their own writing. However, Lamberg does concede that "the use of self-provided feedback reduces but does not eliminate the students' dependence on their readers. Except for private or highly personal expression, students write for an audience and need the audience's responses" (p. 65). Thus, Lamberg also suggests that "teachers use a combination of self- and peer-provided feedback" (p. 67) throughout the semester.

### **Common procedures for self-evaluation**

As with peer evaluation, students may either write their self-evaluation, discuss their papers orally with a peer group, or do both. Of concern here is how teachers can help students look at their own writing more objectively through the use of checklists and self-evaluation questionnaires.

White and Arndt concede that perhaps a major reason why students have difficulty with assessing their own work is that they take that to mean correcting mistakes in spelling, punctuation, grammar, vocabulary, word choice, etc., rather than focusing on content and organization. White and Arndt give examples of "the sorts of questions writers should ask themselves as they assess how coherently they have presented the information and ideas in their texts" (p. 117). These questions come before, during, and after the student writes. White and

Arndt provide these examples, which they also suggest are helpful in student/teacher conferences. Some of these include:

Before writing

1. What are you going to be writing about?
2. How are you going to put that down on paper?
3. How did you go about choosing your subject?
4. What problems might you run into?

While writing

1. How is it going?
2. What are you writing about now?
3. Where are you now in your draft?
4. What will you do with this piece of writing when it is all done?

After writing

1. How did you go about this?
2. Did you make any changes?
3. What are you going to do next with this piece of writing?
4. What do you think of this piece of writing? (p. 132).

Although these are good basic guiding questions, they do not ask writers to address certain parts of their papers. However, there are many composition textbooks that use more specific self-evaluation checklists as part of writing assignments to guide students through the writing process. For example, Leki (1989), just as she suggests for peer evaluation, provides students with task-specific self-evaluation questions for each paper. For example, Leki asks students to identify and restate their thesis statement, give an outline of their main supporting details, and summarize their conclusions. In addition, they have to identify specific sections of their papers that need more detail, better organization, and so forth. All of these tasks are more specific than asking students to identify their general opinions about their paper; "What did you like the best/the least?"

Lamberg's procedure is that the teacher decides the assignment objectives and criteria that papers must meet. Then, as they write their papers, students follow a checklist or assignment sheet which asks them to focus on certain things to include in their papers. After they have written a draft, students share their work with their peer-response groups and use the same assignment checklist mentioned above to evaluate their peers' papers. Then, students revise their papers based on peer comments. Finally, the teacher reads and evaluates papers using the same checklist before asking the student to revise again. Also, the teacher looks at "the effectiveness of the students in evaluating their own and their peers' writing" (p. 68). The benefits of using checklists are that they clarify the assignment, guide students in reading, measuring, and revising their own papers, and guide the peer-response activity (p. 68). Peer response provides students with larger and facilitates the development of "critical reading skills by responding to each other's papers" (p. 68). By including self-evaluation in this whole process, Lamberg believes that "students will learn to apply their critical skills to their own writing and develop independence and confidence in their abilities to direct and evaluate their own performance" (p. 68). All students are active participants in two-way exchange when using both peer and self-response.

### **Student Perspectives of Evaluation**

Although the primary purpose of this study is to look at how accurate and specific students are when giving feedback to their peers and to themselves, Mangelsdorf (1992) makes a case that teachers should also be concerned with how students perceive the evaluation process, which is an aspect covered in the data collection for this thesis. She noted in her own review of the literature on peer evaluation that peer review is highly recommended by ESL composition

teachers, but little has been said about whether students think peer evaluation is helpful. Her study asked students to respond in writing to these questions:

1. Do you find it useful to have your classmates read your papers and give suggestions for revision?
2. What kinds of suggestions do you often receive from your classmates?
3. What kinds of suggestions are most helpful to you?
4. In general, do you find the peer-review process valuable? (pp. 275-276).

Mangelsdorf also asked teachers to write "what they thought were the strengths and weaknesses of the peer review technique" (p. 277). She found that both the students' positive and negative responses were similar to the teachers'. Some positive points that they noted: peer review helps writers clarify their points, exposes writers to a "diversity of thought", helps improve organization and transitions, and helps writers to be more critical. Some negative points: it is difficult to trust other non-native speakers' judgments about correctness; classmates are apathetic; it is discomfoting to criticize peers (p. 277).

Mangelsdorf also found that teachers and students disagreed in one important area, one that is important to all ESL composition teachers and students: "the effects that peer reviews had on student reviewers" (p. 279). The teachers in her study felt that students were made more responsible for their work because of the peer review process and that they gained more confidence about their critical reading and writing ability. In sharp contrast, students "believed that peer reviews had neither helped them to be responsible for their improvement, nor be confident in their ability to critique a text. The problem was that these students did not think that they, or their peers, could be good critics" (p. 280).

Because of this discrepancy between teacher and student perspectives, Mangelsdorf claims that it is not enough to train students in the process of peer

review; it is also necessary that teachers be aware of how students view this process and how to improve it.

Mangelsdorf did not ask her subjects to address directly whether they felt peer review could help a person become a better self-reviewer; nevertheless, her study provides a good background for the follow-up questionnaire on student perspectives of peer and self-evaluation used in this thesis.

### Chapter Summary

This chapter has attempted to outline underlying principles from research in cooperative learning and the process approach to writing for using peer and self-evaluation in L2 composition. Although there are differences between L1 and L2 writers in how peer and self-review work, the research mentioned above shows that both peer and self-evaluation are important activities for helping L2 students become more effective writers and communicators. By looking at different facets of these processes, teachers can determine where they may further improve peer and self-evaluation for their students. There are no guarantees that any kind of feedback will be taken into consideration by students as they revise papers; some student writers may even ignore feedback entirely. But by receiving a variety of input on their papers (from self, peer, and teacher), students should realize that other people have valid suggestions that can help their papers. Particularly, students must learn to provide accurate and specific feedback in order for evaluation to be worthwhile.

The next chapter, Method and Materials, will detail the research design chosen for this study. In addition, the chapter will discuss the composition classes that were chosen to participate in this study, the instrumentation used to collect data, and the procedure that was followed.

### CHAPTER 3. METHOD AND MATERIALS

This chapter gives an explanation of the research design chosen for this study and also discusses the setting for the study, the subjects' backgrounds, and the instrumentation and procedure used to collect the data for this thesis.

#### Rationale for the Research Design

The design of this study is what Seliger and Shohamy (1989) call descriptive. This approach was chosen because the nature of this research is not experimental; that is, it does not "attempt to control or manipulate any of the factors in the research environment" (p. 133). Rather, this study started with a preconceived focus and specific research questions to be explored and described within the context of a pre-existing composition class. Also, although Seliger and Shohamy place both qualitative and descriptive research in the same category, they note that, unlike qualitative research, descriptive research can be analytic in its approach and deductive in its research objectives. Thus, descriptive research shares characteristics with both experimental/quantitative and qualitative approaches:

It is similar to qualitative research because it deals with naturally occurring phenomena, using data which may either be collected first hand or taken from already existing data sources such as data from other studies, student records, and so on. It differs from qualitative research in that it is often deductive rather than heuristic, and begins with preconceived hypotheses and a narrower scope of investigation. In this respect, it shares some of the qualities of experimental research. In addition, descriptive research is often quantitative (p. 124).

One major difference between descriptive and experimental research that should be noted is that "in descriptive research no manipulation of naturally occurring

phenomena occurs, while in experimental research, manipulation and control become important measures of both internal and external validity" (p. 118).

A descriptive approach for this thesis was chosen because the purpose of the research is to describe the peer and self-evaluation abilities of individuals in already-formed groups, without being intrusive on the classroom process. Furthermore, this study did not compare a control group with an experimental group. Rather, the researcher and the instructor of the course were concerned with how the individual students within the group were or were not able to implement the principles and procedure of peer and self-evaluation--the former, a task that had been taught and practiced prior to the research conducted for this thesis, and the latter, a skill that was assumed to have developed as a result of practicing that task.

A descriptive approach also allows for a few generalizations to be drawn from the results of the research because the demographic characteristics of the subjects are somewhat representative of their general population. In this case, the subjects are typical of international undergraduates in Iowa State University composition courses.

Finally, one main factor in choosing a descriptive rather than an experimental approach comes from what Seliger and Shohamy say about the latter:

there has been a growing concern in second language research about the interactive or distorting effects of the research setting on the kind of language data collected (Tarone 1982). Experimental settings, being controlled and artificial, may elicit data different from those produced in natural settings (p. 119).

Thus, it was important to the researcher of this study to select an approach that

would allow her to look at a "real-life" classroom situation and to be as non-intrusive as possible. This thesis also attempts to be analytic and deductive in its explanation of the data collected.

### **Background and Objectives for English 101C**

Each fall, Iowa State University admits between 150 and 280 international undergraduate students (IUs) who are nonnative speakers of English. These students are not United States citizens, nor do they have permanent residency, so they must show evidence of English proficiency in order to be fully admitted to the university. International undergraduates can show this by having a minimum score of 500 on the Test of English as a Foreign Language (TOEFL) or by holding a degree from a college or university in the U.S., Canada, Great Britain, or Australia. Upon admission to the university, IUs are further tested for their English proficiency and for placement into required undergraduate composition courses. In order to do this, the English as a Second Language (ESL) faculty at Iowa State administer an English Placement Test (EPT), which is also used as a diagnostic tool to identify language deficiencies which may cause IUs problems in their academic work. The test has two parts: a thirty-minute composition and a seventy-five-minute section covering listening comprehension, vocabulary, and reading. The writing sample is completed by the students without knowing the topic beforehand and with no extra time to plan, revise, or edit their writing. Since this thesis is concerned with composition only, explanation of the second part of the EPT will not be covered.

Based on their performance on the impromptu writing sample from the EPT, international undergraduates are placed into one of two ESL composition classes, or they pass into the regular freshman English sequence (English

104/105). If IUs do not pass the EPT, they are placed into either English 101B or 101C, which are ESL classes that prepare international students for academic writing in other university courses, including 104/105.

English 101B is an intermediate-level writing course which includes a grammar review and expansion of composition skills such as paragraph development, vocabulary building, and mechanics awareness. Throughout the semester, students in this class prepare to meet the standards for the next level of writing class. Moving from dependence on teacher-provided feedback and evaluation to reliance on their own judgments concerning correctness in both grammar and mechanics is one of the primary goals that students should meet. They are further encouraged to meet this goal in the next course, English 101C, which is the writing class chosen for this study.

English 101C is an advanced-level ESL composition course for international undergraduate students who have either completed 101B or been placed into the class based on the results of their EPT essay. Objectives for English 101C are detailed on pages four and five of the Iowa State University ESL instructor's manual. These include:

1. Introducing students to some of the types of writing they will need to produce in the university, such as compositions similar to those written in English 104 and summaries of lectures or articles;
2. Providing students with the opportunity to read and discuss various types of academic writing;
3. Preparing students to meet the basic standards of correctness in 104, which are set by the Freshman English Committee;
4. Helping students move from dependence on teacher-provided feedback to reliance on self-provided feedback concerning content and form; and

5. Providing discussion of various linguistic and cultural concepts that students must understand in order to communicate effectively in an American academic setting.

Peer evaluation is commonly used in 101C as a technique for students to become better self-evaluators. For the particular subject groups that participated in this study, peer evaluation was used for the first three paper assignments prior to the assignment from which data were collected for this study.

### **Subjects**

The twenty-nine subjects of this study were international undergraduate students at Iowa State University in two of four English 101C sections taught by the same instructor during the 1993 fall semester. Students came from the following Asian countries: Bangladesh (1), Hong Kong (8), India (1), Indonesia (6), Japan (1), Malaysia (6), Singapore (1), South Korea (2), Taiwan (2), and Vietnam (1). Subjects self-reported that they had studied English from 3 to 17 years.

The research in this study was approved by the Iowa State University Human Subjects Research Committee and all subjects in the study participated voluntarily.

### **Instrumentation**

For data collection, peer and self-evaluation questionnaires approved beforehand by the instructor of the course were administered to the subjects. In accordance with procedures previously designed by the instructor, the items on these questionnaires were selected on the basis of their relevance to the paper assignment and were open-ended, task-specific questions. More specifically, students were asked to give descriptive responses rather than simple "yes" or

"no" answers. Leki (1989), the author of the text used for these sections of 101C, designed the paper assignments to focus on some particular feature of writing; for example, the assignment used in this study focuses on deriving the main ideas from responses to a survey and supporting these main ideas with details and examples. To coincide with the design of the writing assignments, Leki provides suggested task-specific peer and self-evaluation guide questions for each assignment, which the instructor of the course modified to fit her own criteria. Also, because both the 101C teacher and Leki follow a process approach, students were asked to concern themselves primarily with content and organization when responding to the first drafts of each other's papers and not to focus on grammar until subsequent drafts. White and Arndt (1991) also support the practice of responding to a first draft solely in terms of meaning and purpose, even if students eventually are encouraged to edit their peers' drafts for mechanics and style.

The task-specific peer and self-response questions used for this particular paper assignment focused on the criteria and guidelines for the assignment and asked students to evaluate their peers and themselves based on whether they had met the parameters of the assignment. Because different students have different strengths and weaknesses in their writing, it is impossible to control for a certain number of errors each student may have. It is also not feasible to assume that all students will have the same content and organization problems; thus, the questionnaire was designed so that students would at least be asked to look for the same general elements in each other's papers, rather than all possible problems a particular paper may have. For future peer review questionnaires, perhaps one question could be included which asks students to

make any other comments and suggestions that the other questions did not address, but for this study, this type of question was not included.

To understand how the peer and self-evaluation questions were chosen, it is important to look at the guidelines for writing the survey paper that the students were given. (A full copy of the assignment sheet can be found in the Appendix).

All papers were expected to have the following sections and information:

- I. Background
  - A. Previous research (if applicable)
  - B. Why are you conducting this survey?
  - C. Hypothesis - What information are you trying to elicit?
  
- II. Description
  - A. Subjects
    1. How many people participated?
    2. Demographics (sex, age, where they are from...)
    3. Selection procedure (volunteer, random selection...)
  - B. Procedure
    1. What did you do?
    2. How did you do it?
  
- III. Results
  - A. What did you find out?
  - B. Use your table to show data
    1. Numbers
    2. Percentages
  
- IV. Discussion
  - A. Did you find any trends?
  - B. What would you do differently if you were to conduct this survey again?

When developing the evaluation questionnaire for this assignment, these guidelines were considered by the researcher.

The students' main focus for the first drafts of this paper was on developing, supporting, and organizing main ideas (content and organization),

but not the grammar, of their peers' and their own first drafts. Thus, students responded to the following questions:

1. Does the paper have four clearly developed sections: background, description of subjects and procedures, results, and discussion? How do you know? If not, how does the writer need to improve his or her overall organization?
2. In which parts does the writer give clear and adequate details to explain the survey? What parts need more detail and explanation? Please specify particular sentences and/or paragraphs.
3. Did the writer find any trends or patterns in the data? What are they? (Give examples.)
4. If the writer didn't interpret the results very clearly or completely, what suggestions can you give so that the paper is better? How can the writer better support his or her findings?
5. Does the writer include his or her own opinion as well as the objective data? Is it clear that the opinion is separate from facts that were collected? Explain why you think so.

A full copy of the questionnaire can be found in the Appendix.

In order to collect information regarding students' opinions of the peer and self-review processes and their perceptions of their ability to give feedback, students were also asked to answer the following items:

1. Do you feel that having a peer respond to your draft helps you improve your writing? If so, why? Why not?
2. Do you prefer your peer to give you written comments only, verbal comments only, or a combination of the two methods? Why or why not?
3. When you and your classmate compared responses to each other's papers, did you tend to agree or disagree? What specifically in your papers did you agree and/or disagree about?
4. If you disagreed, how did you resolve your differences?

5. Which were you better able to do: finding the strengths and weaknesses in your classmate's essay or finding the strengths and weaknesses in your own?

6. If you feel you were better at one type of editing over the other (for example, your peer editing was better/easier than your self editing), why do you think this is so?

7. Do you think that being a peer editor for someone else's paper helps you to edit your own writing more carefully? Why or why not?

A sample of this questionnaire is also in the Appendix.

### Procedure

The teaching and classroom activities for the survey paper unit took place before students were asked to participate in this study; thus, this section deals primarily with the peer review stage of this process. However, an overview of the total assignment and classroom preparation for this paper will be discussed. The writing assignment used in this study is based on one from the textbook used by the instructor (Academic writing: Techniques and tasks, "Chapter 5. Focusing on main ideas"). For this assignment, students learn to gather information about a topic from surveys, analyze the survey data, and draw conclusions about the topic. In order to prepare students for this assignment, the instructor showed two sample surveys and charts to illustrate what they would need to devise for their own surveys. Students also brainstormed various topics that they could use for these. The instructor allowed students to form their own groups based on what topic they were most interested in, and then they worked together in those groups to write drafts of their surveys and data collection charts. After showing these drafts to other groups for suggestions, students made final copies and distributed them to their subjects. Throughout this process, the teacher also used exercises from the Leki text on extracting main ideas as practice

for what the students would be doing with their own surveys. As a final part of the preliminary activities leading up to the actual paper assignment, the instructor also gave out sample survey write-ups for students to analyze for content and organization. These samples also formed the basis for the specific writing assignment given.

After students gathered responses to their survey questionnaires, they had to look for patterns in the data and draw some conclusions about this information. At this stage of the assignment, students wrote their own survey report drafts using their data from the survey. On the day that the first drafts were due, the students conducted a peer evaluation session, which is where this study begins.

Because students in these particular sections of English 101C had been trained for and had gone through the peer review process on three prior paper assignments, no pilot study was done to determine the clarity of the peer evaluation task for this thesis. The instructor of the course did not routinely have students fill out a self-evaluation questionnaire for the previous assignments. Nevertheless, this study tried to determine if students had become better self-evaluators as a result of becoming stronger peer evaluators.

The format of using written comments, rather than oral discussion, for peer evaluation was chosen to maintain continuity in what had been done in previous classes, and to be consistent with the instructor's goals for the class. On the three previous paper assignments, the instructor had students exchange papers in class and answer peer response questions that were tailored for each paper assignment.

The peer and self-evaluation procedures for this study took two fifty-minute class periods for each section of 101C (Section A and Section B). Classes were conducted as usual without any special treatment. For both Section A and Section B, students were told this was part of a study but that the procedure was to be identical to previous peer review class sessions.

On Day 1 students exchanged papers and wrote comments by answering the questions on the peer response sheet. (See Appendix.) Those questions focused on specific elements of the paper and were specific to the survey report assignment. Both the drafts and the peer evaluation comments on them were coded with numbers (100's for Section A and 200's for Section B) to identify the pairs of papers and to ensure anonymity for the participants.

On Day 2 students got back their own papers and answered the same questions for their own draft that were used for the peer evaluation procedure during the previous class period. They did not see their peer's comments until after they completed their self-evaluation. They were allowed the entire fifty minutes for self-evaluation, but all were finished within thirty to forty minutes. It was then decided by the researcher to allow students to meet with their partner and discuss similarities and differences in their reactions, although this was not directly recorded for purposes of data collection. Students did, however, make comments about this part of the procedure on their follow-up questionnaires.

The follow-up questionnaires were given to the students during the next class period to complete on their own time. All twenty-nine students returned these questionnaires to the instructor during the next two class periods.

As a supplement to the follow-up questionnaires, the 101C instructor provided samples of students' end-of-semester evaluations of the class which

directly addressed the peer evaluation procedure. These comments provide further insight into students' impressions of the peer and self-evaluation, and although these remarks were not quantitatively analyzed, samples are included for discussion.

### **Scoring of Peer and Self-Evaluation Tasks**

Students' peer and self-evaluations were read by the evaluator and scored on an ordinal scale for their accuracy and their specificity. That is, for each evaluation question, student received two scores--one for accuracy and one for specificity--which were combined to form a total evaluation score. The total possible for all five questions was 30: 15 for accuracy and 15 for specificity. Highly accurate responses were given a score of 3, partially accurate responses rated a 2, and inaccurate responses received a 1. Blank answers were identified as 0, just as they would be if they were given in as homework. Scores for specificity also ranged from 0-3: 3 = highly specific, 2 = moderately specific, 1 = vague, and 0 = blank. All designators, such as "highly accurate" and "highly specific", were chosen arbitrarily, but the criteria for these categories were chosen on the basis of how completely and thoroughly students answered the evaluation questions. Examples of student responses to the first evaluation question will be shown here to give a clear picture of the range of accuracy and specificity. A highly accurate peer response correctly identifies whether the writer has met all the criteria of each question. A highly accurate answer also addresses all parts of the evaluation question; thus, if the student only says "yes" or "no" to the first part of the question, but does not elaborate, then the response is judged as only partially accurate. A student's response is judged as an inaccurate answer if he either fails to see the problem or misidentifies something as a problem which

really is not. Also, a highly specific response should inform the writer about how he or she should improve the paper, rather than simply pointing out that there are problems. If the evaluator perceived his peer's paper as having no major problems and thus offered a limited answer (simple "yes" or "no" responses), his response was judged highly accurate and highly specific if the peer's paper did indeed have no major problems. To illustrate the range of responses, the next section provides examples from the students' evaluations and explanation as to why the response was given the score if received.

### Student examples

To get a sense of the range of responses and how they differ in terms of accuracy and specificity, three student responses and their scores will be presented in this section. For convenience, the evaluation question is repeated here: "Does the paper have four clearly developed sections: background, description of subjects and procedures, results, and discussion? How do you know? If not, how does the writer need to improve his or her overall organization?" The original responses are reprinted here without spelling or grammar corrections.

Student A responded, "Discuss a bit about your result (get from the question)." This response received a low accuracy rating of 1 and a low specificity rating of 1. Student B: "Yes, it has 4 clearly developed sections. The writer divided it by subtitles." This received a 2 for both accuracy and specificity. Student C writes, "Yes. There are 6 paragraphs in this essay. 1=background, 2 & 3=description of subjects, 4=results, 5=discussion, 6=conclusion. I think that the discussion paragraph needs more writing. I need to know the opinion of the writer on this survey more." This received a 3 for both accuracy and specificity.

These particular samples were, by chance, from students who received the same score for both accuracy and specificity; however, not all students received identical scores in both accuracy and specificity. Some received a 1 for accuracy, but a 3 for specificity, or vice-versa. Thus, these examples are not necessarily representative of the whole group, but rather are meant to show a range of responses. A table of the subjects' scores is presented in the Appendix.

In the next chapter of this thesis, Results, how the students responded to the research tasks will be detailed and possible reasons for the results will be proposed and explained.

## CHAPTER 4. RESULTS

In this chapter, results are presented and analyzed for the peer and self-evaluation tasks in the following ways: The response accuracy scores for each question on both the peer evaluation (AP) and the self-evaluation (AS) tasks were calculated for each student, and then the mean accuracy scores for each student were analyzed for possible correlation. These means are represented as indexes (IAP and IAS). In statistical terms, an index is a number that is used to identify a piece of information. For example, a person's weight is considered a natural index; it is not arbitrarily chosen. Students' individual mean evaluation scores are also given an index, 1, 2, 3, and so on, which are not natural; they are arbitrary. The IAP and the IAS are indexes which summarize a group of data, in this case, the mean accuracy score for each student on peer and self-evaluation. Analyzing the two indexes IAP and IAS was done to determine any connection between students' peer evaluation accuracy score and their self-evaluation accuracy score. That is, IAP and IAS were analyzed to determine if students were neither more nor less accurate on peer evaluation than on self-evaluation.

Next, the response specificity scores for each question on both the peer evaluation (SP) and the self-evaluation (SS) tasks were calculated for each student, and then the mean specificity scores (ISP and ISS) for each student were analyzed for possible correlations. By analyzing these indexes, the researcher hoped to find a connection between students' peer evaluation specificity scores and their self-evaluation specificity scores. The IAP and ISP, as well as IAS and ISS, were also analyzed for possible correlations. The goal again was to determine any connection between student's peer evaluation accuracy and specificity scores and their self-evaluation accuracy and specificity scores. Finally,

the AP and SP scores for all of the subjects were combined to form a total peer evaluation task score (ITP), and the AS and SS scores for all the subjects were calculated to combine a total self-evaluation task score (ITS). These two indexes were then analyzed for possible correlations. The main goal of analyzing these variables was to determine any connection between students' total peer evaluation score with their total self-evaluation score.

In addition to the total task analyses conducted above, one item (#3) from the evaluation questionnaire was also analyzed for possible correlations; for example, the peer evaluation accuracy scores and the self-evaluation accuracy scores for this question were compared to find a connection. Also, results from the follow-up questionnaire were analyzed to determine if student perceptions of their ability to perform peer and self-evaluation matched their actual performance on these tasks.

One important fact to mention before presenting the results is that the data collected for this study are primarily ordinal; that is, the data are "ordered or ranked according to some hierarchical system, such as test scores" (Seliger and Shohamy, 1989 p. 94). For example, 1 is higher than 0, 2 is higher than 1, and so on. Because the data for this thesis is ordinal, only nonparametric statistics, which are used "when we have data that specify just order or ranks or proportions and not precise observational values" (Sprent, 1989 p. 3), are appropriate. Nonparametric statistics are often not strong enough to definitively reject null hypotheses, but they are valuable in showing possible correlations between variables. Thus, measures of frequency and probability were used with the data. Each measure will be explained later in this chapter.

Because some subjects turned in only incomplete sets of data (e.g., only peer evaluations, but not self-evaluations), the data set is quite small in size. Thus, it was determined from the data collected and analyzed that there is not enough evidence of dependence between the peer and self-evaluation tasks to draw any definite conclusions about whether students are more proficient at one task over the other. However, there are observable tendencies in the findings, which will be covered in Chapter 5.

All computer statistical analyses reported below were calculated using SAS (Statistical Analysis System), version 6.7.

### **Results of Peer and Self-Evaluation Tasks**

This section reports the results of both evaluation tasks. First, the results for response accuracy will be presented, followed by the results for response specificity. Finally, the results for the total task (accuracy and specificity scores combined) will be presented.

#### **Explanation of statistical analyses**

The frequencies of peer and self-evaluation scores were analyzed for rank order by using the following statistics: The Chi-square and the Fisher's exact test (2-tail) were used to determine probability of dependence between the accuracy and specificity scores for peer evaluation and self-evaluation; the Kendall's tau-b and the Spearman rank correlation coefficient were used to determine correlations between the two tasks. Each of these measures can be used to determine rank order and correlations with bivariate data, such as the scores of the two evaluation tasks.

Chi-squares test the differences between observed and expected frequencies and are used to determine distributions of independent variables (e.g., AP, AS). For this thesis, significance is at the 5% level. "Generally speaking, high values of chi-squared indicate significance" (Sprenst, 1989 p. 228); thus, these calculations were analyzed for the data collected here. However, due to missing information from some of the 29 subjects for some of the tasks, the computer analysis warns that chi-square may not be an appropriate measure for this study. Specifically, in the case of the chi-squares calculated here, some of the expected cell frequencies were less than five, making this measurement a poor estimate of the actual probability of dependence between variables.

Since the chi-square calculations are not valid here, the Fisher's exact test is used. This tests for independence between variables in two-way tables and "is applicable for inferences conditional on fixed marginal totals" (p. 172). It yields the probability of actually observed data being associated with expected frequencies. Significance is at 5% for this thesis. The closer the probability is to zero, the more likely the variables are dependent upon each other.

The Spearman rank correlation coefficient is appropriate for calculating correlations for variables on an ordinal scale. That is, by using this measurement, we can determine if the rank order for the peer evaluation scores is related to the rank order of the self-evaluation scores. The critical value of significance depends on the number of subjects and is compared to the correlation coefficient. (The critical value can be found in tables at the end of almost any basic statistics text. For this study, Sprenst, 1989 and Gibbons, 1985 were consulted.) If the value calculated for the two variables is close to zero, then the variables are independent and thus not correlated, and if the coefficient

is greater than the critical value, then there is evidence of dependence between variables, thus implying a correlation.

Finally, Kendall's tau-b looks for concordance between pairs of observations, but unlike the Spearman correlation, it accounts for ties in rank. Kendall's tau-b is appropriate only for ordinal variables; therefore, it is an appropriate measure for the data collected here. Like the Spearman correlation, if the coefficient is greater than the critical value, then there is evidence of dependence between variables, thus implying a correlation. (The critical value was found in tables provided by Sprent, 1989 and Gibbons, 1985.) If the value calculated for the two variables is close to zero, then the variables are independent and thus not correlated.

Both the Spearman correlation and the Kendall's tau-b have an ASE (asymptotic relative efficiency) measure, which is used for comparing two tests (Sprent, 1989 p. 11). It is based on the notion that the power or efficiency of one test must be relative to that of another in the same situation. For example, the probability of error on one test should match the probability for the other. Another way to state this is that ASE evaluates "the relative merits of two or more comparable test statistics" (Gibbons, 1985 pp. 18-19). As with this study, "especially in the case of small samples . . . the implications" of the ASE value "cannot be considered particularly meaningful" (p. 19). Thus, this study doesn't look at the ASE value for the Kendall and the Spearman measures, but it does look at the critical value of significance.

### **Accuracy results**

To investigate Hypothesis 1 (H1), "Students are neither more nor less accurate in providing feedback to themselves than to their peers," it was

important to determine any correlation between IAP and IAS. If there is a correlation between the two variables, then H1 can be rejected with some certainty. Frequencies of the scores of the students' responses were calculated first and are presented in Table 1.

Table 1. Frequencies for IAP by IAS.

IAP/IAS	1.4	1.6	1.8	2.0	2.2	2.4	Total
Frequency							
1.2	0	0	0	0	0	0	0
1.4	0	1	0	0	0	0	1
1.6	1	0	0	0	0	0	1
1.8	1	0	1	1	1	0	4
2.0	0	4	0	2	1	0	7
2.2	0	0	1	1	0	0	2
2.4	0	1	0	1	0	1	3
3.0	0	0	0	0	1	0	1
Total	2	6	2	5	3	1	19

Frequency missing=10

The content of this table (and all frequency tables included within this chapter) represents all of the mean scores received by all students who completed the AP and AS tasks. The top row (1.4, 1.6, 1.8, and so on) are the actual indexes observed in the data for AS. The first column (1.2, 1.4, 1.6, and so on) are the actual indexes observed for AP. The data in rows 1-8, columns 2-7 mark the number of students receiving those scores. Finally, the last row and last column show the total number of subjects responding to AP and AS. In this case, only 19 subjects submitted complete peer and self-evaluations. Ten students failed to submit either a peer or a self-evaluation. When these results are plotted on a scattergram, there appears to be no correlation between IAP and IAS. A copy of the scatterplot for IAP by IAS appears in the Appendix.

In order to determine if these frequencies are correlated, thus showing a relationship between the two tasks (AP and AS), and supporting the H1, the statistics in Table 2 and Table 3 were calculated. Results of the Fisher's exact test show that the probability of IAP and IAS being dependent upon each other is 0.404. Since this is greater than 0.050, there is not enough evidence of dependence between the two sets of scores, which disallows for any definite conclusions to be made about H1. Further statistical analyses support this. The

Table 2. Chi-square and Fisher's Exact Test (2-Tail) for IAP by IAS

Statistic	DF	Value	Prob
Chi-square	30.00	32.903	0.327
Fisher's Exact Test (2-Tail)			0.404

Table 3. Kendall's Tau-b and Spearman Rank Correlation Coefficient for IAP by IAS

Statistic	Value	ASE
Kendall's tau-b	0.349	0.172
Spearman correlation	0.428	0.206

value for Kendall's tau-b (see Table 3) is only slightly greater than the critical value of significance (0.333) and shows a only a slight positive correlation between IAP and IAS. The Spearman correlation (see Table 3) shows no connection ( $0.428 < 0.460$ ). Thus, H1 is not rejected.

### Specificity results

In order to address Hypothesis 2 (H2), "Students are neither more nor less specific when responding to task-specific questions for themselves than when doing so for their peers," scores for ISP and ISS were calculated, and the

Table 4. Frequencies for ISP by ISS.

ISP/ISS										
Freq.	1.0	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	Total
1.0	0	0	0	0	0	0	0	0	0	0
1.2	0	0	0	0	0	0	1	0	0	1
1.4	0	0	0	0	1	0	0	0	0	1
1.6	0	0	0	0	0	0	1	0	0	1
1.8	0	1	0	0	0	0	0	1	0	2
2.0	0	0	1	1	1	0	2	0	0	5
2.2	0	0	2	0	0	0	0	0	0	2
2.4	0	0	1	0	1	0	0	0	0	2
2.6	0	0	0	0	0	1	0	1	0	3
2.8	0	0	1	0	0	0	0	0	0	1
3.0	0	0	0	0	0	1	0	0	0	1
Total	0	1	5	1	3	2	4	2	1	19

Frequency missing=10

number of students receiving each score were tabulated (Table 4). The rows and columns have the same desinators as Table 1. That is, the top row indicates the scores for ISS and the first column shows the scores for ISP. The remaining rows and columns show the number of individual subjects receiving the individual scores. A scattergram, included in the Appendix, shows that these frequencies do not appear to be positively or negatively correlated.

Table 5 and Table 6 show the statistics for ISP by ISS. The result of the Fisher's exact test (Probability = 0.598) is greater than 0.050, thus indicating that ISP and ISS are independent of each other. (See Table 5.) Further statistics also show this to be true. Kendall's tau-b shows a value that is close to zero (-0.026) and is less than the critical value of significance (-0.333), thus showing no evidence of correlation. Spearman's correlation also gives a value that is close to zero (-0.041) and is less than the critical value of significance (-0.460), also showing no evidence of correlation. Therefore, H2 is not rejected.

Table 5. Chi-square and Fisher's Exact Test (2-Tail) for ISP by ISS

Statistic	DF	Value	Prob
Chi-square	63.00	59.343	0.607
Fisher's Exact Test (2-Tail)			0.598

Table 6. Kendall's Tau-b and Spearman Rank Correlation Coefficient for ISP by ISS

Statistic	Value	ASE
Kendall's tau-b	-0.029	0.204
Spearman correlation	-0.041	0.242

### Comparison of accuracy and specificity results

This section provides results which address Hypotheses 3a and 3b (H3a, H3b), "Students are neither more accurate nor more specific when conducting peer evaluation. Students are neither more accurate nor more specific when conducting and self-evaluation." First, H3a is analyzed. The frequencies of IAP by ISP were calculated and are shown in Table 7. Again, this table shows the number of individuals who received the actual observed scores for IAP and ISP.

A scattergram of this information shows no correlation between the two variables and is included in the Appendix. The statistics showing probability and correlations appear in Table 8 and Table 9. The Fisher's exact test gives a probability value of 0.336, which is greater than 0.050, and indicates that IAP and ISP are not dependent upon each other. The value for Kendall's tau-b is 0.348, which is greater than the critical value (0.287), and the value for the Spearman correlation is 0.424, which is greater than the critical value (0.400) With these measures, there appears to be slight correlation between IAP and ISP, but, since the probability is larger than 0.050, these values do not reject H3a.

Table 7. Frequencies of IAP by ISP

IAP/ISP	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	Tot.
1.2	1	0	0	0	0	0	0	0	0	0	0	1
1.4	0	0	0	1	0	0	0	0	0	0	0	0
1.6	0	0	0	0	0	0	2	0	0	0	0	2
1.8	0	0	0	0	1	2	0	0	0	1	0	4
2.0	0	0	1	0	2	2	1	1	1	0	0	8
2.2	0	0	0	0	0	1	2	1	0	0	0	4
2.4	0	0	1	0	0	0	0	1	1	0	1	4
3.0	0	0	0	0	0	0	0	0	1	0	0	1
Tot.	1	1	1	1	0	5	5	3	3	1	1	25

Frequency missing=4

Table 8. Chi-square and Fisher's Exact Test (2-Tail) for IAP by ISP

Statistic	DF	Value	Prob
Chi-square	70.00	94.167	0.029
Fisher's Exact Test (2-Tail)			0.336

Table 9. Kendall's Tau-b and Spearman Rank Correlation Coefficient for IAP by ISP

Statistic	Value	ASE
Kendall's tau-b	0.348	0.166
Spearman correlation	0.424	0.196

The relationship between the self-evaluation accuracy scores (IAS) and the self-evaluation specificity (ISS) was determined in order to support or reject H3b. Frequencies of actually observed individual scores were calculated and are shown in Table 10. A scattergram of this information shows no apparent positive or negative correlation between the two variables and appears in the Appendix.

Table 10. Frequencies for IAS by ISS

IAS/ISS										Total
Freq	1.0	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	
1.4	0	1	1	0	0	0	0	0	0	2
1.6	1	0	2	1	0	0	2	1	0	7
1.8	0	0	2	1	0	0	0	0	0	3
2.0	0	1	0	0	3	0	1	1	0	6
2.2	0	0	1	0	0	1	1	0	1	4
2.4	0	0	0	0	0	1	0	0	0	1
Total	1	2	6	2	3	2	4	2	1	23

Frequency missing = 6

Statistics for this information appear in Table 11 and Table 12. The Fisher's exact test gives a probability value of 0.359, which is greater than 0.050. Thus, there is not enough evidence of dependence between these two variables. The value of Kendall's tau-b is 0.329, which is greater than the critical value (0.296), thus showing a slight correlation between IAS and ISS. However, the Spearman correlation is 0.390, which is closer to zero than the critical value (0.416), thus indicating no correlation. Therefore, H3b is not rejected.

Table 11. Chi-square and Fisher's Exact Test (2-Tail) for IAS by ISS

Statistic	DF	Value	Prob
Chi-square	40.00	44.471	0.289
Fisher's Exact Test (2-Tail)			0.359

Table 12. Kendall's Tau-b and Spearman Rank Correlation Coefficient for IAS by ISS

Statistic	Value	ASE
Kendall's tau-b	0.329	0.156
Spearman correlation	0.390	0.191

### Total peer and self-evaluation task results

This section examines H4, "Students are neither better nor worse when conducting peer evaluation overall than they are when conducting self-evaluation overall". The combined mean accuracy and specificity scores were calculated for each student for each task. Then, the mean scores for the whole group were tabulated for both tasks. Frequencies of actual individual scores for ITP by ITS are shown in Table 13.

Table 13. Frequencies for ITP by ITS

ITP/ITS	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.5	Tot.
1.1	0	0	0	0	0	0	0	0	0	0	0	0	0
1.5	0	0	0	0	0	0	0	1	0	0	0	0	1
1.6	0	0	0	0	0	0	0	1	0	0	0	0	1
1.8	0	1	0	0	0	0	0	0	0	0	0	0	1
1.9	0	0	1	0	0	0	0	1	0	1	2	0	5
2.0	0	0	0	1	1	0	0	0	0	0	0	0	2
2.1	0	0	0	1	0	0	0	1	0	0	0	0	2
2.2	0	0	0	0	0	0	0	1	0	0	0	0	1
2.3	0	0	0	0	2	0	0	0	0	0	0	1	3
2.4	0	0	0	0	0	0	0	0	0	0	0	0	0
2.5	0	0	0	0	0	0	0	0	1	0	0	0	1
2.7	0	0	0	0	0	0	0	0	0	0	1	0	1
2.8	0	0	0	0	0	0	0	0	0	1	0	0	1
Tot.	0	1	1	2	3	0	0	5	1	2	3	1	19

Frequency missing=10

In Table 13, actually observed indexes are shown. That is, all actual total scores for both ITP and ITS are given, as well as the number of individual students who received the scores. A scattergram of this information shows no apparent correlation between the two variables and appears in the Appendix.

Statistics for ITP and ITS are given in Table 14 and in Table 15. Results of

the Fisher's exact test show that there is no probability that the two variables are dependent upon each other ( $0.452 > 0.050$ ). Kendall's tau-b and the Spearman correlation both show no correlation between ITP and ITS. The value of Kendall's tau-b shows 0.207, which is less than the critical value of significance (0.333), and the value of the Spearman correlation shows 0.250, which is less than the critical value of significance (0.460). Thus, these variables are considered independent, and H4 is not rejected.

Table 14. Chi-square and Fisher's Exact Test (2-Tail) for ITP by ITS

Statistic	DF	Value	Prob
Chi-square	80.00	87.104	0.275
Fisher's Exact Test (2-Tail)			0.452

Table 15. Kendall's Tau-b and Spearman Rank Correlation Coefficient for ITP by ITS

Statistic	Value	ASE
Kendall's tau-b	0.207	0.156
Spearman correlation	0.250	0.211

### Analysis of Evaluation Question 3

This section looks at the subjects' responses to one particular question (Q3) from the peer and self-evaluation questionnaire: "Did the writer find any trends or patterns in the data? What are they? (Give examples.)" When tallying the accuracy and specificity scores for this question, it appeared that the subjects' accuracy and specificity scores tended to match. That is, if the students received a score of 3 for accuracy, they seemed to receive the same for specificity. Even though the results presented above show that there is no clear correlation

between the peer and evaluation skills of the subjects, it was determined that by analyzing this one question perhaps some conclusions could be drawn about question types that elicit both highly accurate and highly specific responses and, thus, show a correlation between peer and self-evaluation skills. Also, by analyzing this question, possible suggestions for the development of more successful evaluation questionnaire items can be given.

### Accuracy results

Again, the same statistical analyses were used to determine possible correlation between the accuracy scores of the peer responses (Q3AP) and the accuracy scores of the self-responses (Q3AS) for this question. Table 16 shows the frequencies of Q3AP by Q3AS, and a scatterplot (included in the Appendix) indicates that there is a positive correlation between these two variables.

Table 16. Frequencies of Q3AP by Q3AS

Q3AP/Q3AS				
Frequency	1.0	2.0	3.0	Total
1.0	2	1	0	3
2.0	0	3	1	4
3.0	0	3	9	12
Total	2	7	10	19

Frequency missing=10

As with Table 1, this table represents the actual scores received by subjects for Q3AP and Q3AS and the frequency, or number of individual students receiving these scores. To more clearly determine correlations between these two variables, the same statistics used previously were applied here.

Table 17 and Table 18 show the results for this analysis. The Fisher's exact

test shows that the probability of dependence between Q3AP and Q3AS is high:  $0.0046 < 0.0500$ . Results of the Kendall's tau-b and the Spearman correlation corroborate this. The value calculated for the Kendall's tau-b is 0.661, which is greater than the critical value (0.333). This indicates dependence between the variables, thus showing they are correlated. The value of the Spearman correlation (0.687) is also greater than the critical value of significance (0.460),

Table 17. Chi-square and Fisher's Exact Test (2-Tail) for Q3AP by Q3AS

Statistic	DF	Value	Prob
Chi-square	4.0	16.014	0.0030
Fisher's Exact Test (2-Tail)			0.0046

Table 18. Kendall's Tau-b and Spearman Rank Correlation Coefficient for Q3AP by Q3AS

Statistic	Value	ASE
Kendall's tau-b	0.661	0.147
Spearman correlation	0.687	0.151

also indicating a correlation between Q3AP and Q3AS. Upon first investigation, it appears that Q3 elicits accurate responses on both peer and self-evaluation tasks.

### Specificity results

It is also of interest to determine any correlation between the specificity scores for peer and self-evaluation for question 3. Frequencies for Q3SP by Q3SS are shown in Table 19. A scatterplot of this information, included in the Appendix, shows no apparent correlation between the two variables.

Table 19. Frequencies for Q3SP by Q3SS

Q3SP/Q3SS				
Frequency	1.0	2.0	3.0	Total
1.0	0	0	1	1
2.0	0	0	3	3
3.0	1	5	9	15
Total	1	5	13	19

Frequency missing=10

Statistics were calculated to further determine this and are shown in Table 20 and Table 21. The Fisher's exact test shows a low probability that the two variables are dependent upon each other ( $0.723 > 0.050$ ). Further analysis corroborates this. The value of Kendall's tau-b is slightly less than the critical value ( $-0.332 < -0.333$ ), almost showing a negative correlation, but no correlation nonetheless. The value of the Spearman correlation ( $-0.346$ ) is less than the critical value ( $-0.460$ ) and shows no evidence of dependence between the two variables. Therefore, they are not correlated, but they are not too far away from being negatively correlated.

Table 20. Chi-square and Fisher's Exact Test (2-Tail) for Q3SP by Q3SS

Statistic	DF	Value	Prob
Chi-square	4.0	2.338	0.674
Fisher's Exact Test (2-Tail)			0.723

Table 21. Kendall's Tau-b and Spearman Rank Correlation Coefficient for Q3SP by Q3SS

Statistic	Value	ASE
Kendall's tau-b	-0.332	0.095
Spearman correlation	-0.346	0.102

### Comparison of accuracy and specificity results

This section looks at the connection between the accuracy scores and the specificity scores for Q3 on both the peer and the self-evaluation tasks. That is, if students score highly on accuracy for peer evaluation, do they also score highly on specificity for the same task? Also, if they score highly on accuracy for self-evaluation, do they also score highly on specificity for that task? First peer evaluation accuracy and specificity scores are compared.

Table 22 presents the frequencies of actual individual scores for Q3AP by Q3SP. A scatterplot of this information, included in the Appendix, indicates that these are positively correlated.

Table 22. Frequencies of Q3AP by Q3SP

Q3AP/Q3SP				
Frequency	1.0	2.0	3.0	Total
1.0	1	2	2	5
2.0	0	2	4	6
3.0	1	0	13	14
Total	2	4	19	25

Frequency missing=4

Further analysis examines possible correlations between the two variables. The statistics for Q3AP by Q3SP are shown in Table 23 and Table 24. The Fisher's exact test indicates a high probability that the two variables are dependent upon each other, thus showing a positive correlation. The value (0.027) is less than the critical value (0.050), thus showing evidence of dependence. Kendall's tau-b and the Spearman correlation also show evidence of dependence between Q3SP and Q3SS. The value of Kendall's tau-b (0.434) is greater than the critical value of significance (0.287), thus indicating a positive

Table 23. Chi-square and Fisher's Exact Test (2-Tail) for Q3AP by Q3SP

Statistic	DF	Value	Prob
Chi-square	4.0	8.004	0.091
Fisher's Exact Test (2-Tail)			0.027

Table 24. Kendall's Tau-b and Spearman Rank Correlation Coefficient for Q3AP by Q3SP

Statistic	Value	ASE
Kendall's tau-b	0.434	0.178
Spearman correlation	0.458	0.188

correlation between the variables. The value of the Spearman correlation (0.458) is also greater than the critical value (0.398); therefore, the two variables appear to be positively correlated.

It is also important to look at the relationship between self-evaluation accuracy and specificity scores. Table 25 shows the frequency of Q3AS scores by Q3SS scores for individual students. A scatterplot of this information, included in the Appendix, shows evidence of a positive correlation between the two variables.

Table 25. Frequencies of Q3AS by Q3SS

Q3AS/Q3SS				
Frequency	1.0	2.0	3.0	Total
1.0	0	2	1	3
2.0	2	3	3	8
3.0	0	1	10	11
Total	2	6	14	22

Frequency missing=7

Statistics for these variables are presented in Table 26 and Table 27. The Fisher's exact test gives a strong probability that the variables are dependent upon each other ( $0.045 < 0.050$ ), thus showing a positive correlation. Further analysis suggests this as well. The value of Kendall's tau-b (0.485) is greater than the critical value of significance (0.307) and shows evidence of correlation between Q3AS and Q3SS. The value of the Spearman correlation also shows evidence of a positive correlation ( $0.536 > 0.425$ ).

Table 26. Chi-square and Fisher's Exact Test (2-Tail) for Q3AS by Q3SS

Statistic	DF	Value	Prob
Chi-square	4.0	9.425	0.051
Fisher's Exact Test (2-Tail)			0.045

Table 27. Kendall's Tau-b and Spearman Rank Correlation Coefficient for Q3AS by Q3SS

Statistic	Value	ASE
Kendall's tau-b	0.485	0.142
Spearman correlation	0.536	0.157

### Follow-up Questionnaire Results

Twenty-nine subjects responded to a follow-up questionnaire regarding their opinions and perceptions of the peer and self-evaluation processes. For this thesis, the responses to questions 5 and 6 of that questionnaire are analyzed and compared to the students' actual performance on the peer and self-evaluation tasks. Analysis of this data set is to support or reject Hypotheses 5a and 5b (H5a, H5b): "Students' perceptions of their ability to conduct peer evaluation are neither higher or lower than their actual performance on this task. Students' perceptions of their ability to conduct self-evaluation are neither higher or lower

than their actual performance on this task." This analysis is primarily descriptive, but includes samples of students' opinions and impressions of the review process and is thus also partially qualitative in nature.

### Student perceptions and actual performance

For convenience, questions 5 and 6 are given here:

5. Which were you better able to do: finding the strengths and weaknesses in your classmate's essay or finding the strengths and weaknesses in your own?
6. If you feel you were better at one type of editing over the other (for example, your peer editing was better / easier than your self-editing), why do you think this was so?

The subjects' responses were determined to be either positive or negative statements about their peer evaluation ability and either positive or negative statements about their self-evaluation ability. Some students felt they were good at both, and some felt they were not good at either. Table 28 and Table 29 show the results of the student responses.

Table 28. Results of follow-up questionnaire: Peer evaluation self-perception

Peer evaluation self-perception (N=29)	Percentage of N
Positive self-perception	86%
Negative self-perception	14%

Table 29. Results of follow-up questionnaire: Self-evaluation self-perception

Self-evaluation self-perception (N=29)	Percentage of N
Positive self-perception	21%
Negative self-perception	79%

A higher percentage of subjects (86%) gave direct positive comments about their perceived ability to conduct peer evaluation than they did for their perceived ability to conduct self-evaluation (21%). A smaller percentage (only 14%) of the subjects made direct negative comments about their ability to conduct peer evaluation, while a larger percentage (79%) indicated they felt they were not helped by self-evaluation. These findings match their actual performance. The group average score for the total peer evaluation task showed that students were moderately accurate and moderately specific (ITP = 2.056). However, the group average score for the ITS was also moderate, but slightly less (1.904). To determine if these scores are significantly different, a t-test procedure and a univariate procedure were calculated.

The t-test compares the means of two groups (Seliger and Shohamy, 1989 p. 231) and was used in this study to determine if any differences found are by chance. Table 30 shows there is no significant difference between the ITP and the ITS averages for the whole group (t-test). The t-test results show that the value of  $\text{Prob} > |T|$  is greater than the critical value of significance ( $0.134 > 0.050$ ); so, considering all cases with a response, there is no evidence that ITP and ITS are different. Therefore, we cannot say that the whole group performed better on ITP than on ITS.

Table 30. T-test for ITP and ITS scores (considering all cases with a response)

	ITP	ITS	T	Prob >  T
N	25.00	23.00		
Mean	2.056	1.904		
S.D.	0.365	0.318		
Variances			1.528	0.134

However, since not all students provided both peer and self-evaluation, it was necessary to determine any difference in ITP and ITS scores for each complete pair of responses from the subjects (univariate procedure). This information is shown in Table 31.

Table 31. Univariate procedure for ITP and ITS scores (considering cases with pairs)

N	Mean	S. D.	T:Mean=0	Prob >  T
19.00	0.137	0.393	1.517	0.147

The univariate procedure results show that the value of Prob > |T| is greater than the critical value of significance ( $0.147 > 0.050$ ); so, considering cases of subjects with pairs of responses, there is not enough evidence that these scores are different. Thus, we cannot say that individual students performed better on ITP than on ITS.

Thus, the last hypotheses (H5a, H5b) had split results. H5a was not rejected because students perceptions matched their actual performance: they were neither higher nor lower. H5b was rejected because the students' self-perceptions of their ability to conduct self-evaluation were lower than their actual ability. Although students' perceptions that they are better at peer evaluation matches their peer evaluation performance because their ITP showed moderate accuracy and specificity, students' perceptions that they were not effective self-evaluators does not match their performance because their ITS score is also moderately accurate and specific. There was no evidence that the ITP and ITS means were significantly different; that is, the values of ITP and ITS were virtually the same. This shows that students were moderately successful at

both types of evaluation. Thus, students are somewhat effective at providing themselves feedback, but they do not realize their ability. However, the fact that there is no significant difference between ITP and ITS makes it difficult to draw any definite conclusions about H5a and H5b.

### Sample student responses

From reading these questionnaires, one gets the sense that students find self-evaluation not at all helpful; in fact, some students indicated that it may be impossible to look at one's own paper objectively. Some students believe, however, that both peer and self-evaluation are valuable. To get a stronger sense of what the students thought about the evaluation process, excerpts from their follow-up questionnaires and end-of-semester class evaluations are included here:

- "Perhaps it's a human nature that we're better in finding other people's mistakes than ours"
- "It's difficult to find weakness of my own's because it is too close to me...it's hard to read my own essay as an audience"
- "[Evaluating] my classmate's paper is easier because it is not [my] opinion, not my paper"
- "I tend to be too absorb in my writing that I disregard other facts. Therefore self-editing doesn't helps me at all"
- "It's easier to spot people mistakes than to see our own mistakes"

These students found that both peer and self-evaluation were beneficial:

- "Finding the strengths and weaknesses in my classmate's essay [was easier]. But by reading his/her essay, I realised my weaknesses, e.g. I will realise that writing a good beginning in a paper is very important and when I got a chance to write my paper. . . I will take care of the beginning."
- "Both are really necessary. It is a pleasure to find a classmate's strengths and weaknesses because I can let them know what is wrong. They will improve on their paper even more than don't have a peer group. It is the same as me. Based on their comments, I realised what are my weakness parts..."

These responses are quite revealing and are in line with what Mangelsdorf (1992) found in her study, which was cited in Chapter 2: students do not always realize their ability to conduct peer and self-evaluation. Perhaps this is because this particular group had not received direct instruction and practice in conducting self-evaluation of their own writing. These reactions to peer and self-evaluation will be discussed more thoroughly in Chapter 5.

### Chapter Summary

Before discussing possible conclusions for these results, it is important to summarize what was discovered in the data analysis. H1, H2, H3a, H3b, and H4 were not rejected because not enough evidence of dependence between variables existed. Because no correlations were supported, it could not be claimed that the variables were related; therefore, the hypotheses were rejected. To be precise, if it is claimed that students are better at peer evaluation than self-evaluation, a connection or probability of dependence between the two tasks must be shown. We cannot say that students develop and improve their self-evaluation skills as a result of having developed and improved their peer evaluation skills. If there is no evidence of correlation, then it cannot be definitely shown that one is better than the other because they are not related in any way. Therefore, we cannot say, for example, that students are more or less accurate on peer evaluation tasks than self-evaluation tasks.

Again, H5a was not rejected because students' perceptions matched their actual performance, but H5b was rejected because the students' self-perceptions of their ability to conduct self-evaluation were lower than their actual ability. There was no evidence that the ITP and ITS means were significantly different; thereby making it difficult to draw any definite conclusions about H5a and H5b.

To further analyze and summarize the results found in this section, Chapter 5 presents possible reasons explaining the results that were calculated here. Also, implications for L2 composition teachers and suggestions for further research will be presented.

## CHAPTER 5. CONCLUSIONS

This final chapter will summarize possible answers to the research questions and hypotheses in the introduction. For convenience, the hypotheses are restated here.

- H1: Students are neither more nor less accurate in providing feedback to themselves than to their peers.
- H2: Students are neither more nor less specific when responding to task-specific questions for themselves than when doing so for their peers.
- H3a: Students are neither more accurate nor more specific when conducting peer evaluation.
- H3b: Students are neither more accurate nor more specific when conducting self-evaluation.
- H4: Students are neither better nor worse when conducting peer evaluation overall than they are when conducting self-evaluation overall.
- H5a: Students' perceptions of their ability to conduct peer evaluation are neither higher nor lower their actual performance on this task.
- H5b: Students' perceptions of their ability to conduct self-evaluation are neither higher or lower their actual performance on this task.

This chapter also presents implications for L2 composition teachers and suggestions for future research.

### Overall Conclusions

This thesis set out to determine if students were able to provide accurate and specific feedback on both peer and self-evaluation tasks, and whether students' self-evaluation skills seemed to develop along with their peer

evaluation skills. While no definitive statements may be made from this small study about these issues, there are some important points to consider.

With the cases of H1 and H2, it was determined that not enough evidence of dependence existed to show that students are no more or no less accurate and are no more or no less specific when providing peer and self-evaluation. This fact is significant, because this means that these hypotheses are not rejected by the data. Although students seem to have similar abilities on the two tasks, it cannot be claimed that skills for one develop from those of the other because they are not related. Also, because of the lack of correlation among the above variables, it might be assumed that H4 would also not be supported. Indeed, when the total peer and self-evaluation tasks (accuracy and specificity scores combined) were analyzed, no relationship between the total scores was found. The claim stated by several sources in Chapter 2 that students' self-response skills become better over time as a result of their becoming better peer evaluators has not been supported by research in the past, and it is not supported by this study. Moreover, becoming proficient and accurate in peer evaluation does not appear to preclude students' developing their self-monitoring abilities. Thus, the point made by Leki (1990) and Stanley (1992) that students need training in conducting evaluation seems to apply to both peer and self-response. While much attention has been given to how to train student writers to conduct peer evaluation, more attention should be given to the teaching and training of students to review their own writing as critically and objectively as possible. Self-evaluation appears to be a separate but necessary process for instructors to teach student writers.

With H3a and H3b, there was not enough evidence of dependence to show that students' accuracy scores are related to their specificity scores for both peer

and self-evaluation. For example, when comparing the accuracy of the subjects' responses on the peer evaluation task to the specificity responses on the same task, we find that whether students are accurate does not depend on whether they are also specific, and vice versa. Some students may have received a 3 for accuracy for certain items on the evaluation questionnaire but only a 1 for specificity for those same items. Because of this disparity, there seems to be no connection between the two skills. Reasons for this are unclear, but perhaps students focus on one element of the response process over the other. For example, a student may accurately identify a problem in her peer's essay, but not be able to give a specific suggestion for correcting the problem. As Leki (1990) noted in her work, students often mimic the vague responses they themselves have received from their peers and teachers, and thus do not realize that they must also provide specific suggestions. In addition to the lack of correlation between accuracy and specificity in peer response, there is also no correlation between these variables in self-response. Again, although there is no apparent connection between peer and self-evaluation skills in this study, it is interesting to note that the same lack of correlation between accuracy and specificity exists when students provide themselves with feedback as it does for peer feedback.

The above results show that this study cannot make any claims that students are better at one type of evaluation (peer) over another (self) because they apparently have no relationship to or bearing on each other; thus, the null hypotheses stand. The results are also not surprising, considering that the subjects who participated in the study had not received direct instruction in self-evaluation. In fact, the results suggest that self-evaluation is a skill that needs to be directly taught and emphasized; it cannot be assumed that students will

develop self-review strategies on their own. Perhaps self-evaluation involves a different thought process than peer evaluation. As many students commented in their follow-up questionnaires, looking at one's own paper as an outside reader is difficult and artificial. Writers are absorbed in their own agendas for writing, thus missing alternative ideas that others might suggest. Unfortunately, product-centered approaches still abound in both L1 and L2 composition, and student writers do not always have the opportunity to receive peer feedback. Therefore, they must be able to look at their own writing critically. This means that self-evaluation is important, and teaching students to evaluate their own writing is necessary no matter how difficult or artificial it seems. Treating self-evaluation as a valuable tool to help writers step back from their papers should be taught and encouraged along with peer evaluation. After all, it is a skill that can be useful in other academic courses where writing is required as well as in the workplace after graduation.

Although the first four hypotheses were not rejected, it was determined that there was enough evidence of dependence between variables of one evaluation question (Q3), which was examined in isolation. Subjects scored highly on both accuracy and specificity for both peer and self-evaluation, which indicated possible correlation among the variables of AP, AS, SP, and SS. Because the subject population is small, conclusions are limited but possible. The main conclusion drawn from these results is that the question type and structure seems to elicit both accurate and specific responses on both peer and self-evaluation tasks. While item analysis is not the main focus of this thesis, it is necessary to note that Q3 asks a direct information ("yes"/"no") question first, followed by a "WH" question: "Did the writer find any trends or patterns in the

data? What are they? (Give examples)", whereas some of the other evaluation questions asked WH questions and elicited disparate responses (accurate but not specific, or specific but not accurate). As Celce-Murcia and Larsen-Freeman (1983) discuss in their text, information questions are the easiest and most direct questions for students to comprehend and answer because they require basic information, "yes" or "no". However, teachers want students to provide more than this type of response--they want students to elaborate--so information questions are often avoided for evaluation purposes. However, the combination of this type of question with "WH" follow-up questions appears to be more effective.

With H5a, there appears to be a relationship between how students perceive their ability to conduct peer evaluation and their actual performance on that task. With H5b, however, students do not seem to be aware of their ability to conduct self-evaluation. The majority viewed self-evaluation as unhelpful, and in fact, they viewed negatively their ability to conduct self-response. However, their self-evaluation performance was equal to their peer evaluation performance and showed moderate accuracy and specificity, which indicates that students do provide themselves with moderately adequate feedback. Therefore, students are apparently able to provide themselves with valid feedback, but they are not aware of their ability to do so.

### **Implications for Composition Teachers**

It is hoped that the results of this research have some significance for a larger audience beyond the primary readers. Thus, some implications for teachers using evaluation in L2 classes will be suggested here. The main issues to consider involve the following areas: training, practice, guidance, and

awareness for students; and teacher consideration of peer and self-response evaluation questionnaire design.

Training students to conduct peer and self-evaluation is a complex but valuable process. It is not enough to ask students to read someone's paper and tell that person what they thought. Students must have a clear set of criteria for assignments and knowledge about the conventions of writing by which to evaluate the work of their peers and their own (Rothschild and Klingenberg, 1990; Stanley, 1992). Accurate and specific responses are requisite if evaluation is to be beneficial.

In addition to learning what is expected of them during the evaluation stage of the writing process, students must have opportunities to practice what they have learned. The subjects for this study had practiced evaluation on models at the beginning of the semester and conducted peer evaluation on previous papers. While their early evaluations were not compared to their later ones for this study, it is hoped that their evaluation skills improved with practice over time. Their overall scores for both peer and self-evaluation were moderately accurate and specific, indicating that they were somewhat skilled in providing feedback by the time this study was conducted. However, this study cannot draw any conclusions about whether this was due to training and/or practice over time.

Whatever the case, teachers need to guide their students during peer review so that students can be aware of appropriate evaluative comments. If a student is giving inconsistent, or simply wrong, information to his peers, the teacher must intervene and determine where the misunderstanding lies. This can be done if teachers either sit in on oral discussion groups, or if they collect

and comment on written evaluation questionnaires. In addition to students learning to give appropriate feedback, teachers must also evaluate their own feedback strategies in order to give students adequate models of responses.

Students must also be aware of their own abilities to conduct evaluation. As with the Mangelsdorf (1992) study, this thesis determined that the majority of subjects viewed self-evaluation negatively when, in fact, they performed moderately well. Awareness of their own ability to self-monitor is crucial for all student writers because, ultimately, they are responsible for a "final" draft of their work and must be accountable for their choices in content, organization, phrasing, and so on. Students also do not always have the opportunity to receive feedback before a paper is due, and thus, must learn to evaluate their own work. Teachers can help students pinpoint their strengths and weaknesses in the skill of self-evaluation in addition to their writing and peer evaluation skills.

Finally, teacher consideration of the evaluation questionnaire design is crucial for peer and self-response to be helpful. Student writers, especially those to whom peer and self-evaluation is a novel concept, need help determining what to look for when critiquing papers. By providing criteria and specific sections for students to examine, teachers can facilitate the peer review process. Since not all student papers are equal in terms of types of strengths and weaknesses, it is important for students to have a set of common items to look for. Although this element was not included on the evaluation questionnaire for this study, it is also beneficial to include at least one open-ended response question at the end of the questionnaire for students to comment on other content and organizational concerns they had with the paper (Mittan 1989). This

less-guided question could satisfy a student's need to provide more personal feedback.

### Suggestions for Further Research

Some changes in future related studies may elicit more definitive conclusions. The primary limitation of this study was the small sample population. If this type of research were repeated in the future, it would be important to control for student attrition by using more sections of similarly-taught composition courses. Gathering data from a larger pool of subjects could perhaps allow for more definite conclusions to be made.

Also, because this study looks at a "snapshot" of the peer and self-evaluation processes, it is impossible to see any connection between the effect of time on the development of students' evaluation skills. For example, the average accuracy score for the whole group on peer evaluation indicates that students are moderately accurate in their responses, but obviously, their comments could be even more precise. It would be interesting to study accuracy levels over time to determine if time and practice are factors in improvement.

To determine the effect of instruction on students' evaluation skills, future research might compare the accuracy and specificity scores of students who have received both peer and self-evaluation training with students who have received no training. Valuable conclusions as to the effectiveness of training may be drawn. (Rothschild and Klingenberg's study addresses this issue.) Cultural or linguistic differences could also be looked for in the patterns of accuracy and specificity.

Another important change to be implemented involves the instructor matching students' papers by level of proficiency for the peer evaluation

procedure. For example, by having a student exchange papers with a peer who has similar strengths and weaknesses in his writing, more definite statements might be made regarding students' abilities to provide accurate and specific feedback on both peer and self-evaluation because they will have worked with papers of the same caliber.

Finally, conducting an item analysis of evaluation questions could provide more information about the type of questions which seem to elicit the most accurate and specific responses from students. The goal would be to determine if task-specific questions really draw more specific suggestions than general evaluation questions. These findings would either reject or confirm claims made that task-specific questions actually elicit specific responses.

The above suggestions are not the only possible changes that could be made in designing further research in the area of peer and self-evaluation. However, these focus on the primary areas of this particular study that could be improved.

Regardless of the limitations of this study, the results suggested that peer and self-evaluation are two separate processes. Students do not necessarily learn to monitor their own work as a result of learning to monitor others' writing. The subjects' own responses to the follow-up questionnaire also indicate that they do not see the validity or utility of self-response. However, students are able to provide themselves varying amounts of feedback and must be encouraged to recognize their self-monitoring abilities. Accuracy and specificity are goals that must be stressed by teachers if evaluation is to be successful, and control over the types of evaluation questions may yield better accuracy and specificity scores.

## REFERENCES

- Applebee, A. N. (1986). Problems in process approaches: Toward a reconceptualization of process instruction. In A. R. Petrosky & D. Bartholomae (Eds.), The teaching of writing: Eighty-fifth yearbook of the National Society for the Study of Education, Part II (pp. 95-113). Chicago: National Society for the Study of Education.
- Ary, D., Jacobs, L.C., & Razavieh, A. (1985). Introduction to research in education (3rd ed.). New York: Holt, Rinehart and Winston.
- Allaei, S., & Connor, U. (1990). Exploring the dynamics of cross-cultural collaboration. The Writing Instructor, 10, 19-28.
- Allen, E. D., & Sutton, T. J. (1964). The effect of practice and evaluation on improvement in written composition. Cooperative Research Project 1992. Deland, Florida: Stetson University. (ERIC Document 001 274)
- Allen, M. S., & Roswell, B. S. (1989). Self-evaluation in holistic assessment. Paper presented at the 40th Annual Meeting of the Conference on College Composition and Communication. Seattle, WA, March 16-18, 1989. 18p. (ERIC Document 303 809)
- Barnes, G. A. (1983). Comments on Vivian Zamel, "Writing: The process of discovering meaning." TESOL Quarterly, 17, 137-138.
- Beaven, M. H. (1977). Individualized goal setting, self-evaluation, and peer evaluation. In C. R. Cooper & L. Odell (Eds.) Evaluating writing: Describing, measuring, judging. (pp. 135-156). Urbana, IL: National Council of Teachers of English.
- Bejarano, Y. (1987). A cooperative small-group methodology in the language classroom. TESOL Quarterly, 21(3), 483-504.
- Bell, J. H. (1991). Using peer response groups in ESL writing classes. TESL Canada Journal/Revue TESL du Canada, 8(2), 65-71.
- Benesch, S. (1985). Improving peer response: Collaboration between teachers and students. Journal of Teaching Writing, 4(1), 87-94.
- Berkenkotter, C. (1983). Student writers and their audiences: Case studies of the revising directions of three college freshmen. Paper presented at the 16th

- Annual Meeting of the Canadian Council of Teachers of English.  
Montreal, Canada, May 10-14, 1983. 14p. (ERIC Document 236 618)
- Berkenkotter, C. (1984). Student writers and their sense of authority over texts. College Composition and Communication, 35(3), 312-319
- Borg, W. R., & Gall, M. D. (1989). Educational research: An introduction (5th ed.). New York: Longman.
- Brown, A. L., & Palinscar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), Knowing, learning, and instruction (pp. 393-451). Hilldale, NJ: Erlbaum.
- Bruffee, K. A. (1984). Collaborative learning and the "Conversation of Mankind." College English, 46(7), 635-652.
- Bruffee, K. A. (1986). Social construction, language, and the authority of knowledge: A bibliographic essay. College English, 48, 773-790.
- Brumfit, C. (1984). Communicative methodology in language teaching: The roles of fluency and accuracy. Cambridge: Cambridge University Press.
- Carson, J., & Nelson, G. (1994). Writing groups: Cross-cultural issues. Journal of Second Language Writing, 3(1), 17-30.
- Celce-Murcia, M., & Larsen-Freeman, D. (1983). The grammar book: An ESL/EFL teacher's course. Boston: Heinle & Heinle.
- Chaudron, C. (1983). Evaluating writing: Effects of feedback on revision. Revised version of a paper presented at the 17th Annual TESOL Convention. Toronto, Ontario, March 16-19, 1983. 26p. (ERIC Document 227 705)
- Christensen, L. (Ed.) (1982). A guide to teaching self/peer editing. Produced as part of the Wisconsin Writing Project with a grant from the National Endowment for the Humanities, Washington D. C. 35p. (ERIC Document 249 513)
- Coe, R. M.. (1987). An apology for form; or, who took the form out of the process? College English, 49(1), 13-28.
- Connor, U. (1987). Research frontiers in writing analysis: TESOL Quarterly, 21(4), 677-696.

- Cooper, C.R., & Odell, L. (Eds.) (1977). Evaluating writing: Describing, measuring, judging. Urbana, IL: National Council of Teachers of English.
- Cumming, A. (1986). Intentional learning as a principle for ESL writing instruction: A case study. TESL Canada Journal/Revue TESL du Canada, Special Issue 1, 69-83.
- Diaz, D. M. (1986a) The adult ESL writer: The process and the context. Paper presented at the 76th Annual Meeting of the National Council of Teachers of English. San Antonio, TX, November 21-26, 1986. 18p. (ERIC Document 281 235)
- Diaz, D. M. (1986b). The writing process and the ESL writer: Reinforcement from second language research. The Writing Instructor, 5(4), 167-175.
- Dicker, S. (1981). Applying the monitor model to the editing of compositions. Paper presented at the 3rd Annual TESOL Summer Meeting. New York, NY, July 24-26, 1981. 17p. (ERIC Document 209 925)
- DiPardo, A., & Freedman, S. W., (1987). Historical overview: groups in the writing classroom. Technical Report No. 4. University of California, Berkeley. Center for the Study of Writing. 21p. (ERIC Document 282 229)
- DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretic foundations and new directions. Review of Educational Research, 58(2), 119-149.
- Elbow, P. (1973). Writing without teachers. New York: Oxford University Press.
- Ewald, H. R. (1984). Evaluation as process, not product. The Writing Instructor, 3(3), 130-138.
- Faigley, L. (1986). Competing theories of process: A critique and a proposal. College English, 48(6), 527-542.
- Flynn, E. (1982). Effects of peer critiquing and model analysis on the quality of biology student laboratory reports. Paper presented at the annual meeting of the National Council of Teachers of English, Washington, DC. (ERIC Document 234 403)
- Flynn, E. A. (1984). Students as readers of their classmates' writing: Some implications for peer critiquing. The Writing Instructor, 3(3), 120-128.

- Ford, B. W. (1973). The effects of peer editing/grading on the grammar-usage and theme composition ability of college freshmen. Ph.D. Dissertation, University of Oklahoma.
- Gere, A. R. (1987). Writing groups: History, theory, and implications. Carbondale: Southern Illinois Press.
- Gere, A. R., & Abbott, R. D. (1985). Talking about writing: The language of writing groups. Research in the Teaching of English, 19(4), 362-385.
- Gere, A. R., & Stevens, R. (1985). The language of writing groups: How oral response shapes revision. In S. W. Freedman (Ed.), The acquisition of written language: Response and revision (pp. 85-105). Norwood, NJ: Ablex.
- George, D. (1984). Working with peer groups in the composition classroom. College Composition and Communication, 35(3), 320-326.
- Gibbons, J. D. (1985). Nonparametric statistical inference, 2nd ed., revised and expanded. New York: Marcel Dekker, Inc.
- Gilbert, D. (1992). Reconsidering the writing workshop. The English Record, 42(3), 1-5.
- Grimm, N. (1986). Improving students' responses to their peer's essays. College Composition and Communication, 37(1), 91-94.
- Hafernik, J. J. (1984). The how and why of peer editing in the ESL writing class. Paper presented at the 1983 CATESOL State Conference, Los Angeles, California. (ERIC Document 253 064)
- Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in teaching of writing. College Composition and Communication, 33, 76-88.
- Halliday, M. A. K., & Hassan, R. (1985). Language, context, and text: Aspects of language in a social-semiotic perspective. Victoria, Australia: Deakin University.
- Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language. Journal of Second Language Writing, 1(3), 255-276.

- Hillocks, G., Jr. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. American Journal of Education, 93, 107-132.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). Applied statistics for the behavioral sciences, 2nd ed. Boston: Houghton Mifflin.
- Hughey, J. B., Wormuth, D. R., Hartfiel, V. F., & Jacobs, H. L. (1983). Teaching ESL composition: Principles and techniques. Rowley, MA: Newbury House.
- Huntley, H. (1992). Feedback strategies in intermediate and advanced second language composition: A discussion of the effects of error correction, peer review, and student-teacher conferences on student writing and performance. 18p. (ERIC Document 355 809)
- Hvitfeldt, C. (1986). Guided peer critique in ESL writing at the college level. Paper presented at the Annual Meeting of the Japan Association of Language Teachers International Conference on Language Teaching and Learning. Seirei Gakuen, Hamamatsu, Japan, November 22-24, 1986. 11p. (ERIC Document 282 438)
- Jacobs, G. (1987). First experiences with peer feedback on compositions: student and teacher reaction. System, 15(3), 325-333.
- Jacobs, G. (1988). Co-operative goal structure: A way to improve group activities. ELT Journal, 42(2), 97-101.
- Jacobs, G., & Ilola, L. M. (1990). Disagreement can be inviting: A cooperative learning approach. Paper presented at the Annual Meeting of the American Educational Research Association. Boston, MA, April 16-20, 1990. 20p. (ERIC Document 319 738)
- Jacobs, G., & Zhang, S. (1989). Peer feedback in second language writing instruction: Boon or bane? Paper presented at the 1989 Annual Meeting of the American Educational Research Association. San Francisco, CA, March 27-31, 1989. 23p. (ERIC Document 306 766)
- Johns, A. M. (1986). Coherence and academic writing: Some definitions and suggestions for teaching. TESOL Quarterly, 20(2), 247-265.

- Johnson, D. W., & Johnson, R. T. (1987). Learning together and alone: Cooperative, competitive, and individualistic learning. Second ed. Englewood Cliffs, New Jersey: Prentice-Hall.
- Kleine, M. (1985). What freshman say--and might say--to each other about their own writing. Journal of Teaching Writing, 4(2), 216-233.
- Kostelnick, C. (1989). Process paradigms in design and composition: Affinities and directions. College Composition and Communication, 40, 267-281.
- Krashen, S. D. (1984). Writing: Research, theory, and applications. Oxford: Pergamon Institute of English.
- Kroll, B. (1978). Sorting out writing problems. In C. H. Blatchford & J. Schachter (Eds.), On TESOL '78: EFL policies, programs, practices (pp. 176-182). Washington, DC: TESOL.
- Kuhn, T. (1970). The structure of scientific revolutions (rev. ed.). Chicago: University of Chicago Press.
- Lagana, J. R. (1974). The development, implementation, and evaluation of a model for teaching composition which utilizes individualized learning and peer group. Ph.D. Dissertation, The University of Michigan.
- Lamberg, W. (1980). Self-provided and peer-provided feedback. College Composition and Communication, 31, 63-69.
- Lapidus-Saltz, W. (1981). The effective feedback script: A peer response procedure. The Writing Instructor, 1(1), 19-25.
- Leki, I. (1989a). Academic writing: Techniques and tasks. New York: St. Martin's Press.
- Leki, I. (1989b). Instructor's manual to accompany Academic writing: Techniques and tasks. New York: St. Martin's Press.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. The CATESOL Journal, 3, 5-19.
- Macrorie, K. (1979). Telling writing. New York: Hayden.
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? ELT Journal, 46(3), 274-284.

- Mangelsdorf, K., & Schlumberger, A. (1992). ESL student response stances in a peer-review task. Journal of Second Language Writing, 1(3), 235-254.
- McKay, S. (1982). A focus on prewriting strategies. In M. Hines & W. Rutherford (Eds.), On TESOL '81 (pp. 89-95). Washington, DC: TESOL.
- McKay, S. (1983). Some limitations in teaching composition. Paper presented at the 17th Annual Convention of Teachers of English to Speakers of Other Languages. Toronto, Canada, March 15-20, 1983. 9p. (ERIC Document 275 151)
- McKendy, T. F. (1990). Legitimizing peer response: A recycling project for placement essays. College Composition and Communication, 41(1), 89-91.
- Millet, N. C. (1969). On snarls and straighteners. Illinois English Bulletin, 57, 13-18.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. M. Johnson & D. H. Roen (Eds.), Richness in Writing: Empowering ESL Students (pp. 207-219). New York & London: Longman.
- Moffett, J. (1968). Teaching the universe of discourse. Boston: Houghton Mifflin.
- Moll, L. C. (1987). Teaching second language students: A Vygotskian perspective. In D. M. Johnson & D. H. Roen (Eds.), Richness in Writing: Empowering ESL Students (pp. 55-69). New York & London: Longman.
- Myers, G. D. (1986). The writing seminar: Broadening peer collaboration in freshman English. The Writing Instructor, 6(1), 48-55.
- Nelson, G. L., & Murphy, J. M. (1992). An L2 writing group: Task and social dimensions. Journal of Second Language Writing, 1(3), 171-193.
- Nelson, G. L., & Murphy, J. M. (1992/1993). Writing groups and the less proficient ESL student. TESOL Journal, 2(2), 23-26.
- Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? TESOL Quarterly, 27(1), 135-141.

- Newkirk, T. (1984a). Direction and misdirection in peer response. College Composition and Communication, 35(3), 301-311.
- Newkirk, T. (1984b). How student read student papers: An exploratory study. Written Communication, 1(3), 283-305.
- Nystrand, M. (1986). Learning to write by talking about writing: A summary of research on intensive peer review in expository writing instruction at the University of Wisconsin-Madison. In M. Nystrand (Ed.), The structure of written communication (pp. 179-211). Orlando, FL: Academic Press.
- Otto, M. K. (1992). Responsive environments, reflective writers: One teacher's goals for change. The English Record, 42(2), 6-10.
- Pierson, H. (1972). Teaching writing. Engelwood Cliffs, NJ: Prentice-Hall.
- Raimes, A. (1979). Problems and teaching strategies in ESL composition (If Johnny has problems, what about Juan, Jean, and Ywe-Han?). Arlington, VA: Center for Applied Linguistics.
- Raimes, A. (1986). Teaching ESL writing: Fitting what we do to what we know. The Writing Instructor, 5(4), 153-166.
- Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. TESOL Quarterly, 25(3), 407-430.
- Reid, J. (1984). Comments on Vivian Zamel's "The composing processes of advanced ESL students: Six case studies." TESOL Quarterly, 18, 149-153.
- Rothschild, D. & Klingenberg, F. (1990). Self and peer evaluation of writing in the interactive ESL classroom: An exploratory study. TESL Canada Journal/Revue TESL du Canada, 8(1), 52-65.
- Rubin, L. (1985). Uneven performance: What students do and don't know about their own writing. The Writing Instructor, 4(4), 157-167..
- Sager, C. (1973). Improving the quality of written composition through pupil use of rating scales. Ph.D. Dissertation, Boston University.
- Santos, T. (1992). Ideology in composition: L1 and ESL. Journal of Second Language Writing, 1, 1-15.

- Schwarte, B., & Matsumura-Lothrop, E. (1987). The self-monitoring of articles and verbs in ESL written production. MinneTESOL Journal, 6, 29-51.
- Seliger, H. W., & Shohamy, E. (1989). Second language research methods. Oxford: Oxford UP.
- Shaughnessy, M. (1977). Errors and expectations. New York and London: Oxford University Press.
- Silva, T., & Zamel, V. (1988). Comments on Vivian Zamel's "Recent research on writing pedagogy": A reader reacts/The author responds. TESOL Quarterly, 22(3), 517-524.
- Slavin, R. E. (1983). Cooperative learning. New York and London: Longman.
- Sommers, J. (1983). Student-teacher memos: A collaborative means to student development. Paper presented at the 16th Annual Meeting of the Canadian Council of Teachers of English. Montreal, Canada, May 10-14, 1983. 20p. (ERIC Document 232 190)
- Spack, R. (1984). Invention strategies and the ESL composition student. TESOL Quarterly, 18, 649-670.
- Spack, R., & Sadow, C. (1982). Student-teacher working journals in ESL freshman composition. TESOL Quarterly, 17, 575-593.
- Sprent, P. (1989). Applied nonparametric statistical methods. London: Chapman and Hall.
- Stanley, J. (1992). Coaching student writers to be effective peer evaluators. Journal of Second Language Writing, 1(3), 217-233.
- Stern, C. (1992). Student writers practicing self-assessment. Paper presented at the 43rd Annual Meeting of the Conference on College Composition and Communication. Cincinnati, OH, March 19-21, 1992. 11p. (ERIC Document 354 509)
- Susser, B. (1994). Process approaches in ESL/EFL writing instruction. Journal of Second Language Writing, 3(1), 31-47.
- Tarone, E. (1982). Systematicity and attention in interlanguage. Language Learning, 32, 69-84.

- Tebo-Messina, M. (1989). Authority and models of the writing workshop: All collaborative learning is not equal. The Writing Instructor, 8(2), 86-92.
- Thornburg, J., & Schmidt, H. H. (1994). Word processing for ESL writers. Workshop presented at the 1994 Computers in Applied Linguistics Conference. Iowa State University, Ames, Iowa, July 8-13, 1994.
- Uruza, C. (1987). "You stopped too soon": Second language children composing and revising. TESOL Quarterly, 21(2), 279-297.
- Villamil, O., & Carrasquillo, A. (1988). Assessing writing in the ESL classroom. 14p. (ERIC Document 301 032)
- Vygotsky, L. (1962). Thought and language. (E. Hanfmann & G. Vakar, Eds. and Trans.). Cambridge, MA: MIT Press.
- Walvoord, B. E. F. (1986). Student response groups: Training for autonomy. The Writing Instructor, 6(1), 39-47.
- Wauters, J. K. (1988). Non-confrontational critiquing pairs: An alternative to verbal peer response groups. The Writing Instructor, 7(3/4), 156-166.
- White, R. & Arndt, V. (1991). Process writing. London: Longman.
- Wiener, H. S. (1986). Collaborative learning in the classroom: A guide to evaluation. College English, 48(1), 52-61.
- Witbeck, M. C. (1976). Peer correction procedures for intermediate and advanced ESL composition lessons. TESOL Quarterly, 10(3), 321-326.
- Wolter, D. R. (1975). Effect of feedback on performance on a creative writing task. Ph.D. Dissertation, The University of Michigan.
- Young, R. (1978). Paradigms and problems: Needed research in rhetorical invention. In C. Cooper & L. Odell (Eds.), Research in composing (p. 31). Urbana, Illinois: National Council of Teachers of English.
- Zamel, V. (1976). Teaching composition in the ESL classroom: What we can learn from research in the teaching of English. TESOL Quarterly, 10, 67-76.
- Zamel, V. (1982). Writing: The process of discovering meaning. TESOL Quarterly, 16, 195-209.

- Zamel, V. (1983a). The composing processes of advanced ESL students: Six case studies. TESOL Quarterly, 17, 165-187.
- Zamel, V. (1983b). Reply to Barnes. TESOL Quarterly, 17, 138-139.
- Zamel, V. (1984a). The author responds. TESOL Quarterly, 18, 154-158.
- Zamel, V. (1984b). In search of the key: Research and practice in composition. In J. Handscombe, R. A. Orem, & B. P. Taylor (Eds.), On TESOL '83 (pp. 195-207). Washington, DC: TESOL.
- Zamel, V. (1985). Responding to student writing. TESOL Quarterly, 19(1), 79-101.
- Zamel, V. (1987). Recent research on writing pedagogy. TESOL Quarterly, 21(4), 697-715.

## ACKNOWLEDGMENTS

I want to express my great appreciation to Dr. Barbara Matthies, my major professor. Her practical guidance and advice helped me to keep the focus of my research. She motivated me to look beyond the confines of my thesis and encouraged me to become a better teacher during the research process. Her direction was invaluable. I also want to express gratitude to the other members of my committee, Dr. Barbara Schwarte and Dr. Fred Duffelmeyer, for their time, suggestions, and most of all, their patience during this process.

Thanks to Dr. Mary Barratt, Dr. Marge Graves, and Lucy Zollner for the use of their office and computer. I greatly appreciated the quiet place to think and write. Also, their help and encouragement gave me the boost I needed to finish.

Special thanks to Phil Bungum for finding Juan Jose Goyeneche and Koji Kondo, who both provided guidance with the statistical analysis of my research data. Also, Molly Lohnes, fellow graduate student and English 101C instructor who willingly provided two sections of her class for this study, deserves recognition.

Finally, I thank my family and my husband, Michael, for their support.

## APPENDIX A. PEER/SELF-EVALUATION QUESTIONNAIRE

## Peer/Self-Response Questions: Survey Report

Please write your responses on a clean sheet of paper and not on the writer's draft. Answer the questions as fully and completely as you can because your responses will count as an in-class assignment.

1 Does the paper have four clearly developed sections: background, description of subjects and procedures, results, and discussion? How do you know? If not, how does the writer need to improve his or her overall organization?

2 In which parts does the writer give clear and adequate details to explain the survey? What parts need more detail and explanation? Please specify particular sentences and/or paragraphs.

3 Did the writer find any trends or patterns in the data? What are they? (Give examples.)

4 If the writer didn't interpret the results very clearly or completely, what suggestions can you give so that the paper is better? How can the writer better support his or her findings?

5 Does the writer include his or her own opinion as well as the objective data? Is it clear that the opinion is separate from facts that were collected? Explain why you think so.

**APPENDIX B. FOLLOW-UP QUESTIONNAIRE****Follow-up Questionnaire for Peer/Self-Evaluation Project**

Please give at least a two sentence response to these questions. Use the back of this form or another sheet of paper if you need more room to write.

- 1 Do you feel that having a peer respond to your draft helps you improve your writing? If so, why? Why not?
  
- 2 Do you prefer your peer to give you written comments only, verbal comments only, or a combination of the two methods? Why or why not?
  
- 3 When you and your classmate compared responses to each other's papers, did you tend to agree or disagree? What specifically in your papers did you agree and/or disagree about?
  
- 4 If you disagreed, how did you resolve your differences?
  
- 5 Which were you better able to do: finding the strengths and weaknesses in your classmate's essay or finding the strengths and weaknesses in your own?
  
- 6 If you feel you were better at one type of editing over the other (for example, your peer editing was better/easier than your self editing), why do you think this is so?
  
- 7 Do you think that being a peer editor for someone else's paper helps you to edit your own writing more carefully? Why or why not?

**APPENDIX C. ASSIGNMENT SHEET FOR SURVEY REPORT****Guidelines for Survey Write-up**

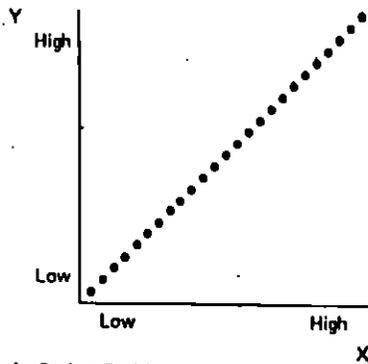
- I. Background
  - A. Previous research (if applicable)
  - B. Why are you conducting this survey?
  - C. Hypothesis - What information are you trying to elicit?
- II. Description
  - A. Subjects
    - 1. How many people participated?
    - 2. Demographics (sex, age, where they are from...)
    - 3. Selection procedure (volunteer, random selection...)
  - B. Procedure
    - 1. What did you do?
    - 2. How did you do it?
- III. Results
  - A. What did you find out?
  - B. Use your table to show data
    - 1. Numbers
    - 2. Percentages
- IV. Discussion
  - A. Did you find any trends?
  - B. What would you do differently if you were to conduct this survey again?

## APPENDIX D. SCORE SUMMARY FOR ALL SUBJECTS

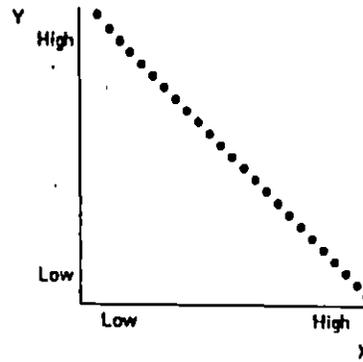
Subject ID	IAP	IAS	ISP	ISS	ITP	ITS	Peer response self-perception	Self-response self-perception
101	1.4	1.6	1.6	2.4	1.5	2.0	positive	negative
102		2.2		1.6		1.9	positive	negative
103	1.8	1.8	2.8	1.6	2.3	1.7	positive	negative
104	2.0	4.6	2.2	1.6	2.1	1.6	positive	negative
105		1.8		1.8		1.8	positive	negative
106	2.4	2.0	1.4	2.0	1.9	2.0	positive	negative
107		1.6		1.0		1.3	positive	negative
108	1.8	1.4	1.8	1.4	1.8	1.4	positive	negative
109		2.0		1.4		1.7	positive	positive
110	2.0		1.8		1.9		positive	negative
111	1.2		1.0		1.1		positive	negative
112	2.4		2.4		2.4		positive	negative
113	1.6		2.2		1.9		positive	negative
201	2.4	1.6	2.6	2.6	2.5	2.1	positive	negative
202	1.6	1.4	2.2	1.6	1.9	1.5	positive	negative
203	2.0	1.6	2.0	1.8	2.0	1.7	positive	negative
204	2.4	2.4	3.0	2.2	2.7	2.3	positive	positive
205	1.8	2.2	2.0	2.4	1.9	2.3	positive	negative
206	3.0	2.2	2.6	2.2	2.8	2.2	positive	negative
207	2.0	1.6	2.0	1.6	2.0	1.6	positive	negative
208	1.8	2.0	2.0	2.4	1.9	2.2	positive	negative
209	2.2		2.2		2.2		positive	negative
210	2.0	1.6	1.2	2.4	1.6	2.0	negative	positive
211	2.0	2.2	2.6	2.8	2.3	2.5	negative	positive
212	2.0	2.0	2.4	2.0	2.2	2.0	positive	negative
213	2.2		2.2		2.2		positive	negative
214	2.0	2.0	1.8	2.6	1.9	2.3	negative	positive
215	2.2	1.8	2.4	1.6	2.3	1.7	positive	negative
216	2.2	2.0	2.0	2.0	2.1	2.0	negative	positive

## APPENDIX E. SCATTERPLOTS OF VARIABLES

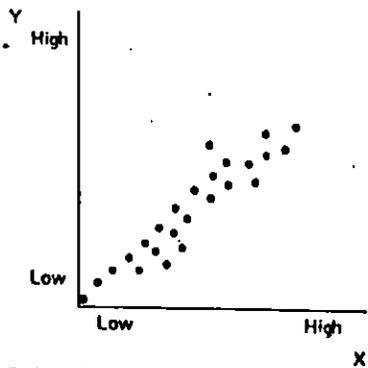
Sample scattergrams from Ary, Jacobs, & Razavieh, 1985 p. 120.



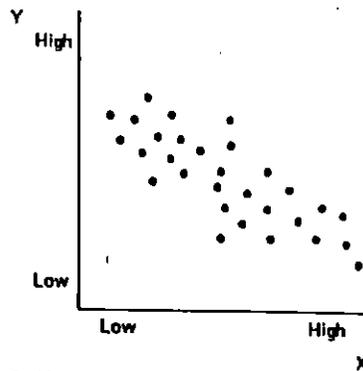
A. Perfect Positive  
Correlation (+1.00)



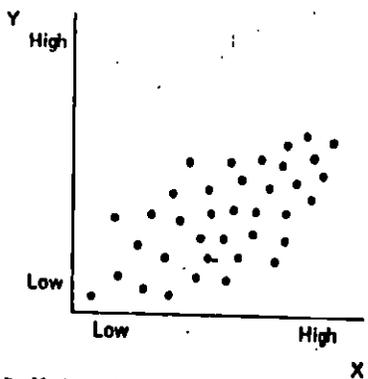
B. Perfect Negative  
Correlation (-1.00)



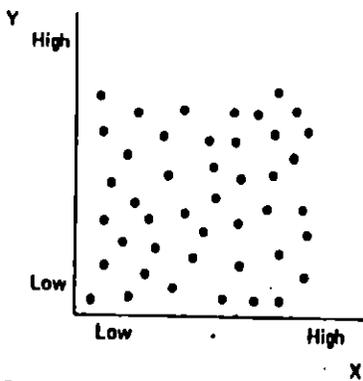
C. High Positive  
Correlation (+.93)



D. Moderate Negative  
Correlation (-.76)

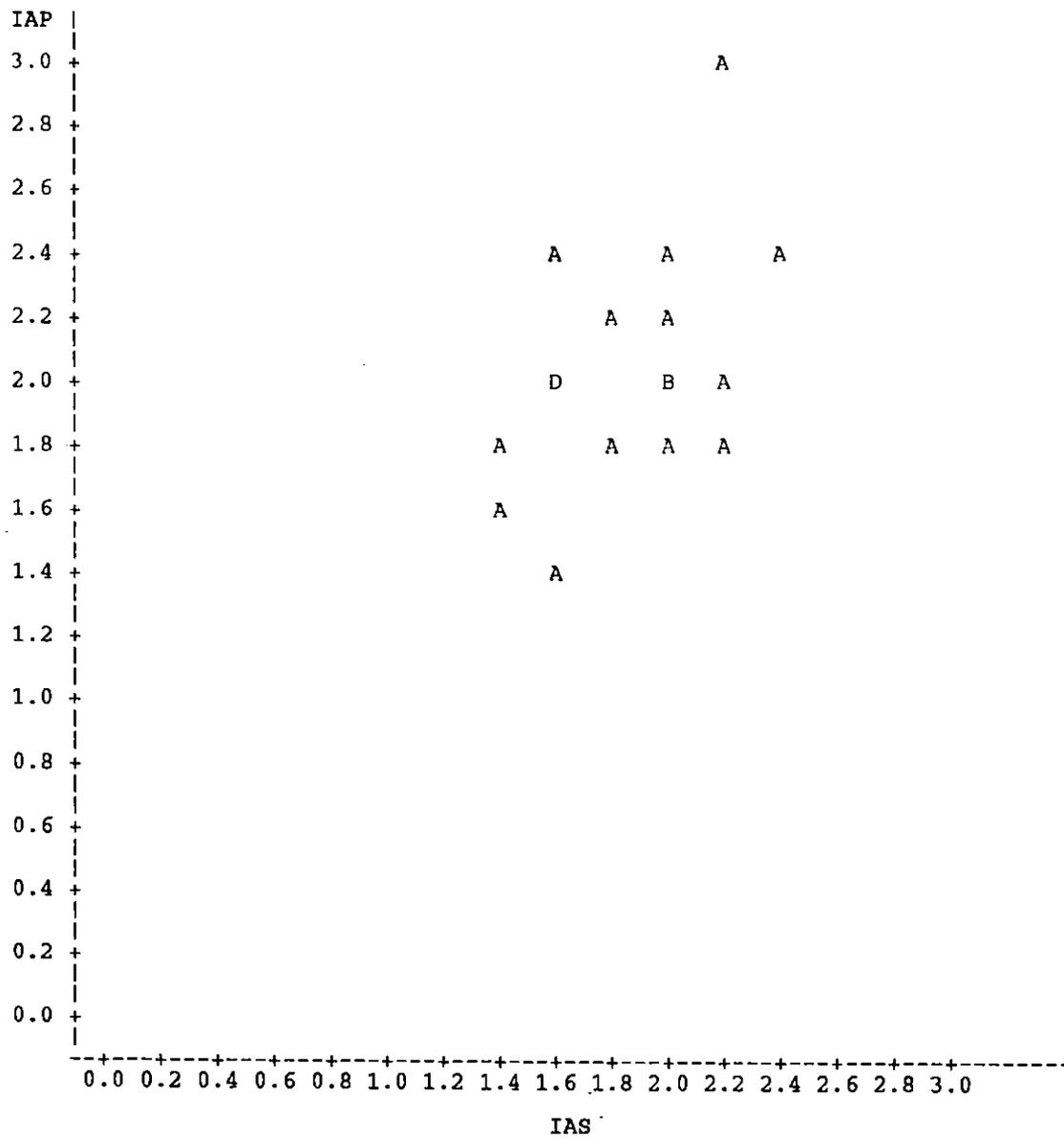


E. Moderate Positive  
Correlation (+.67)

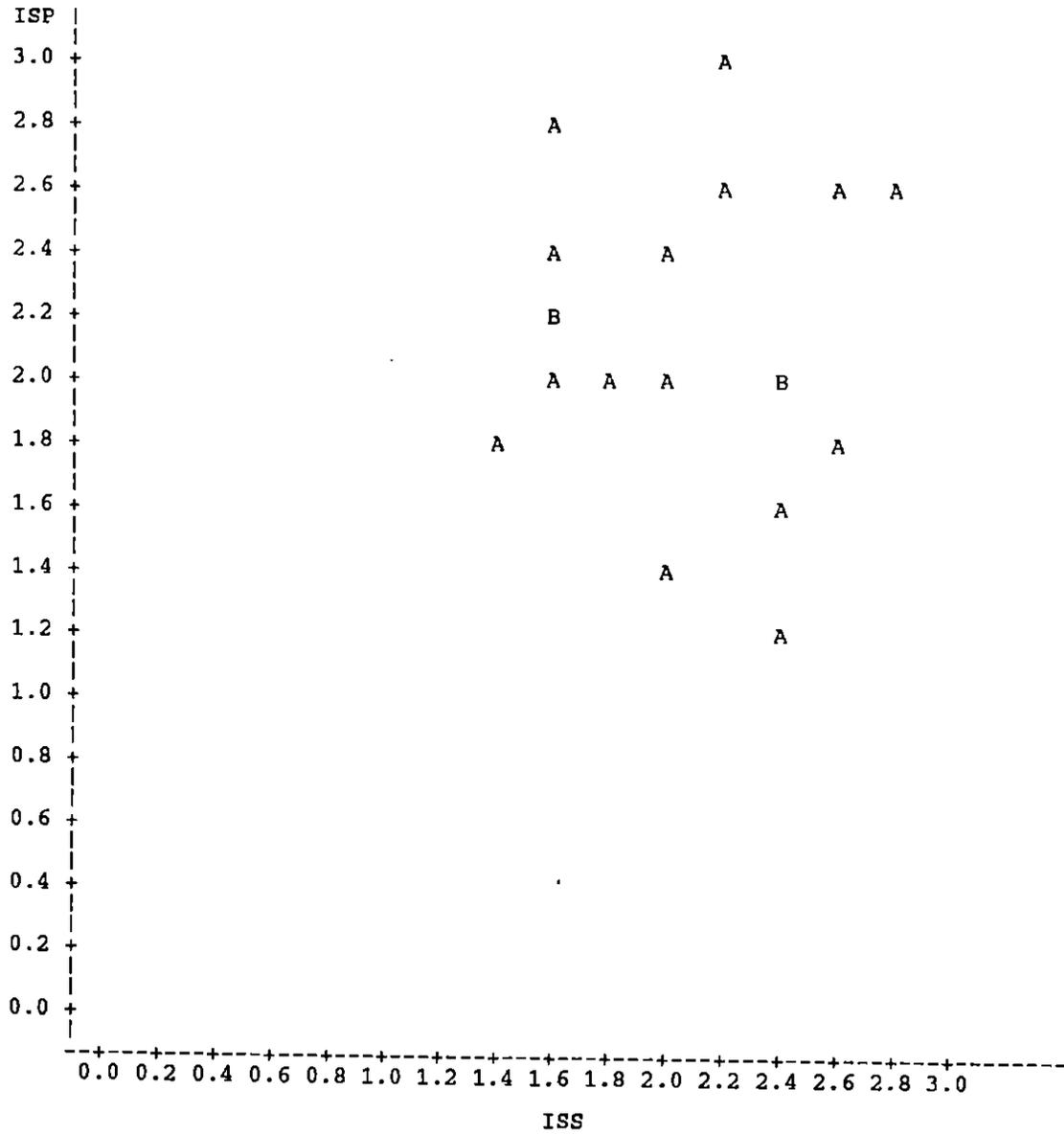


F. Zero Correlation

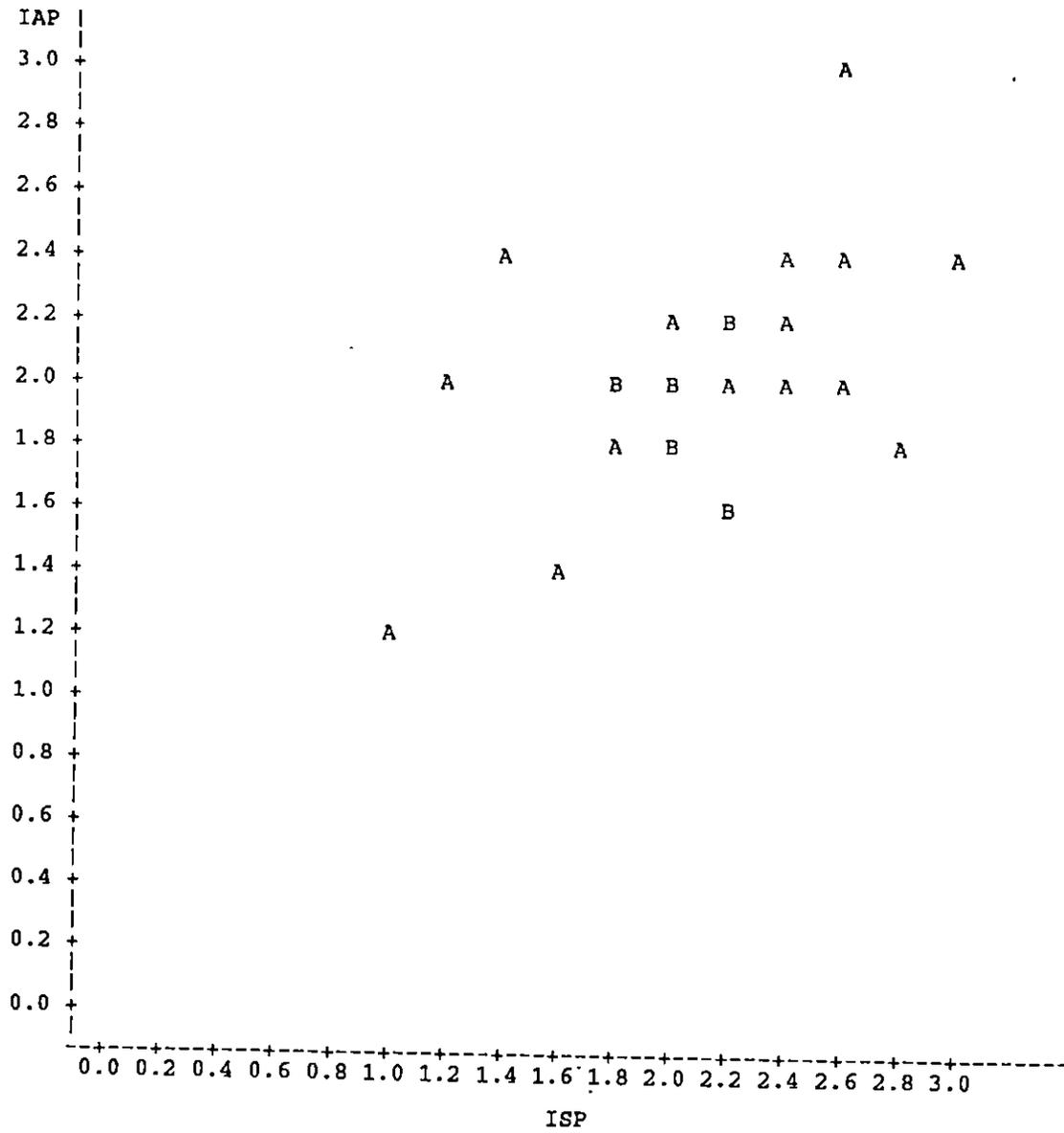
Plot of IAP by IAS. Legend for all plots: A=1 obs, B=2 obs, etc.



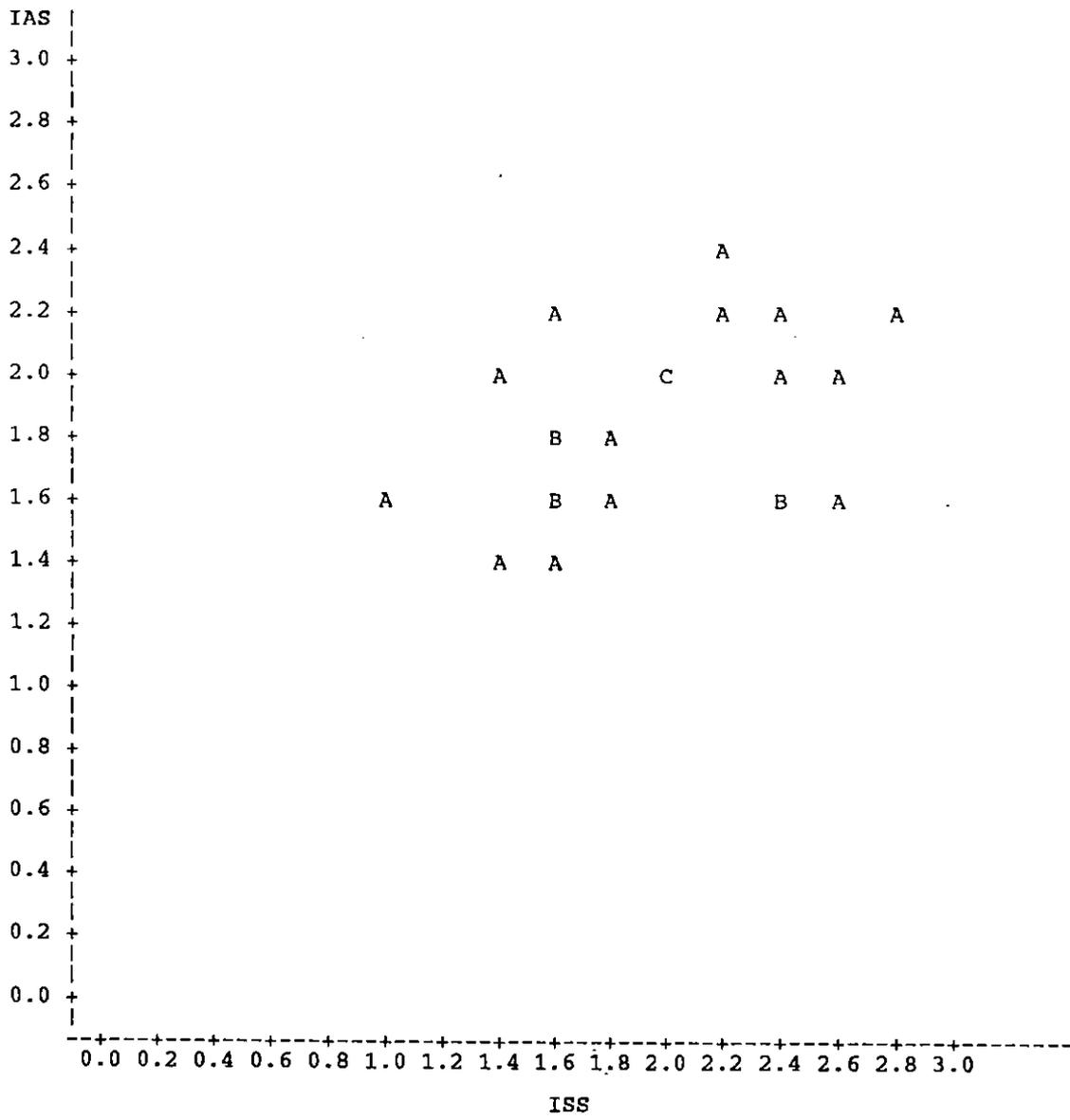
Plot of ISP by ISS.



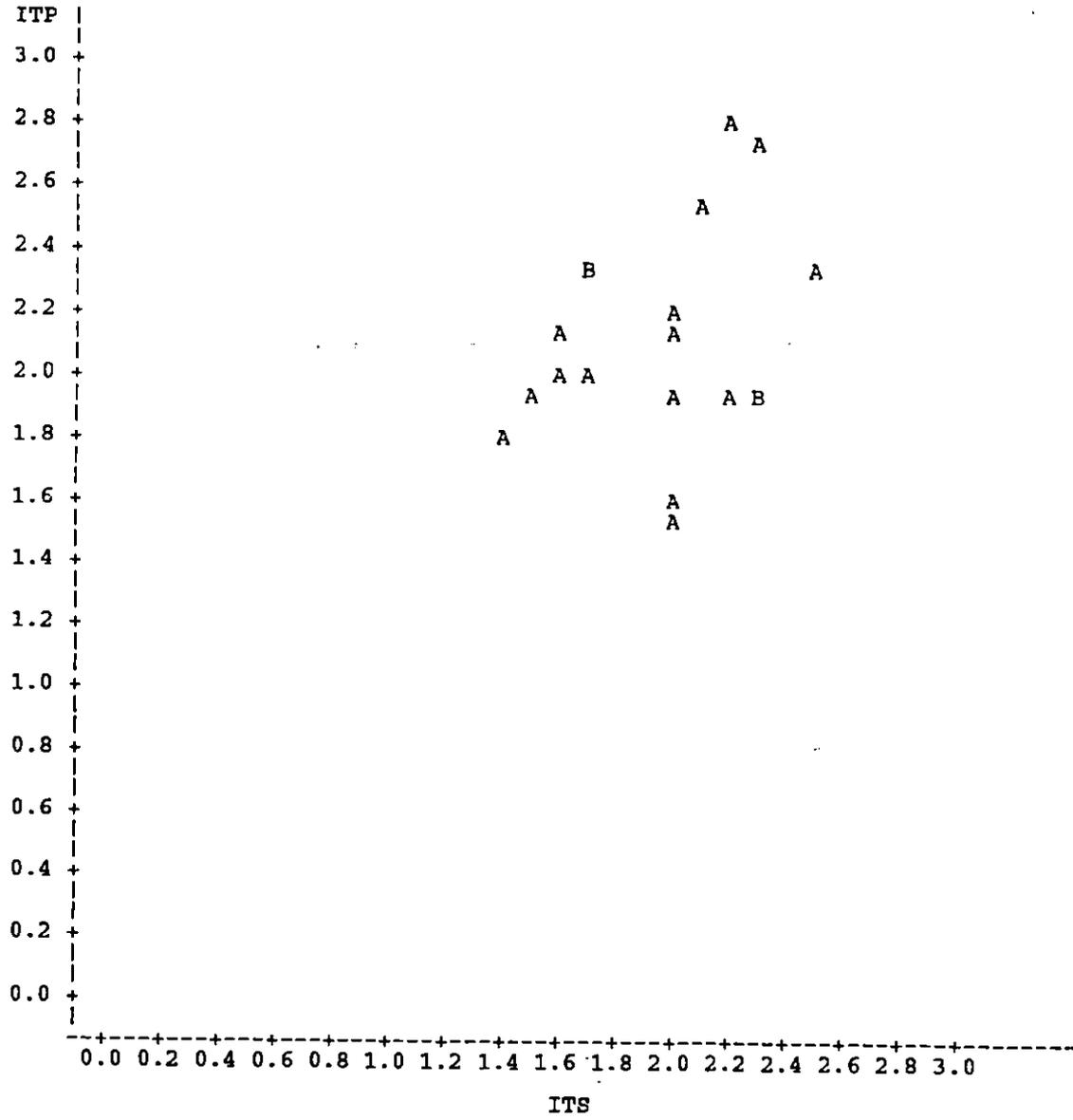
Plot of IAP by ISP.



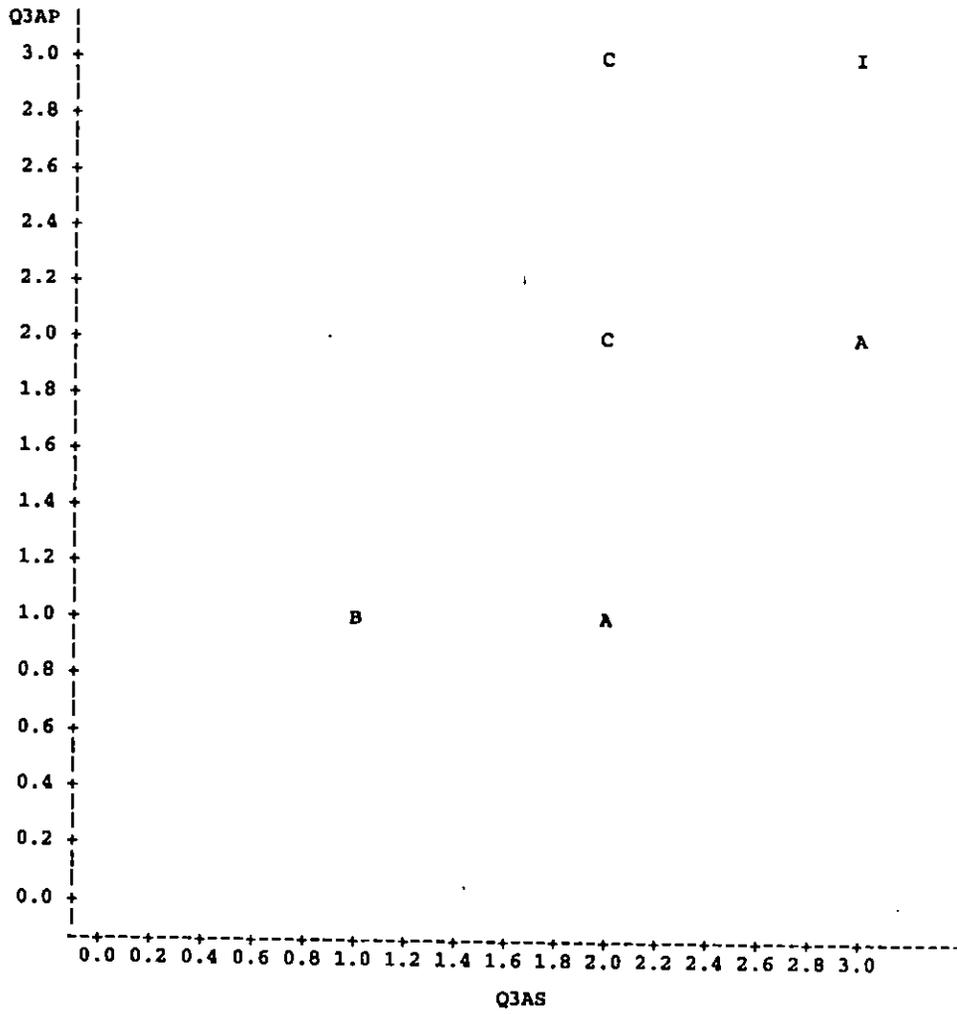
Plot of IAS by ISS.



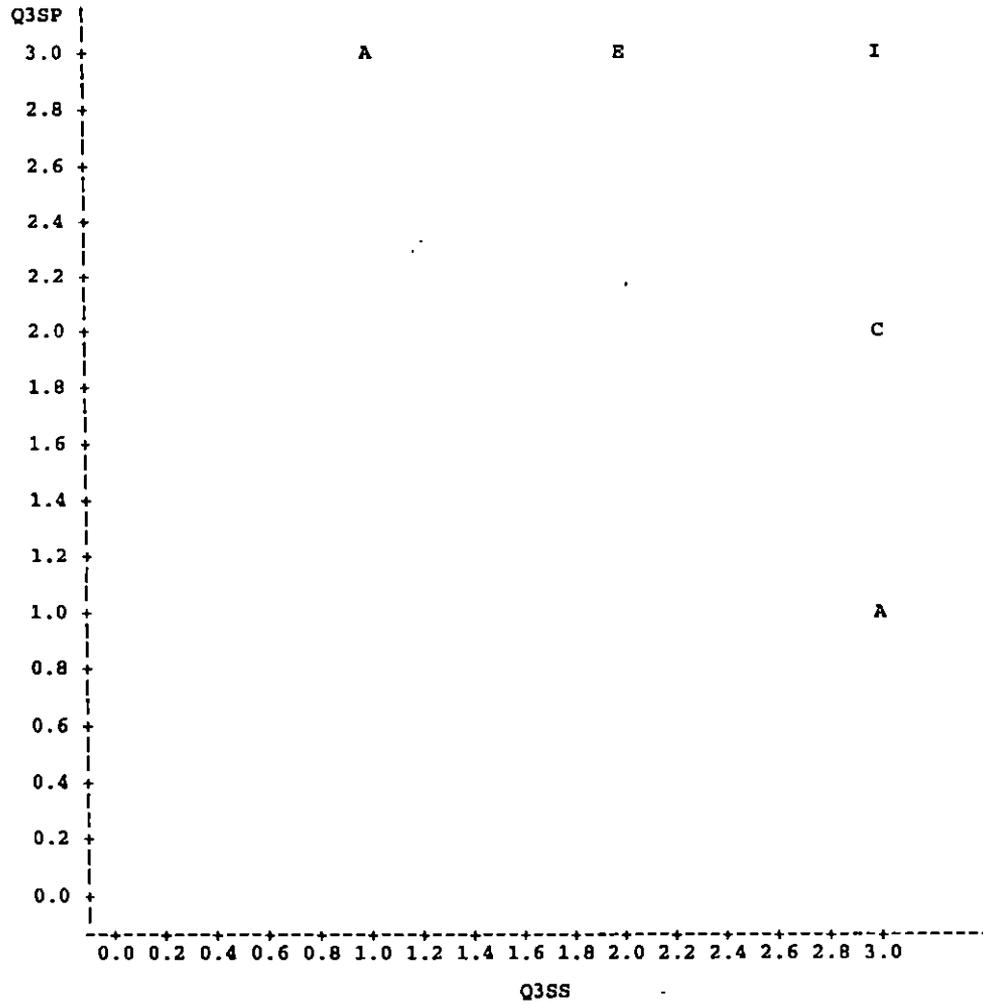
Plot of ITP by ITS.



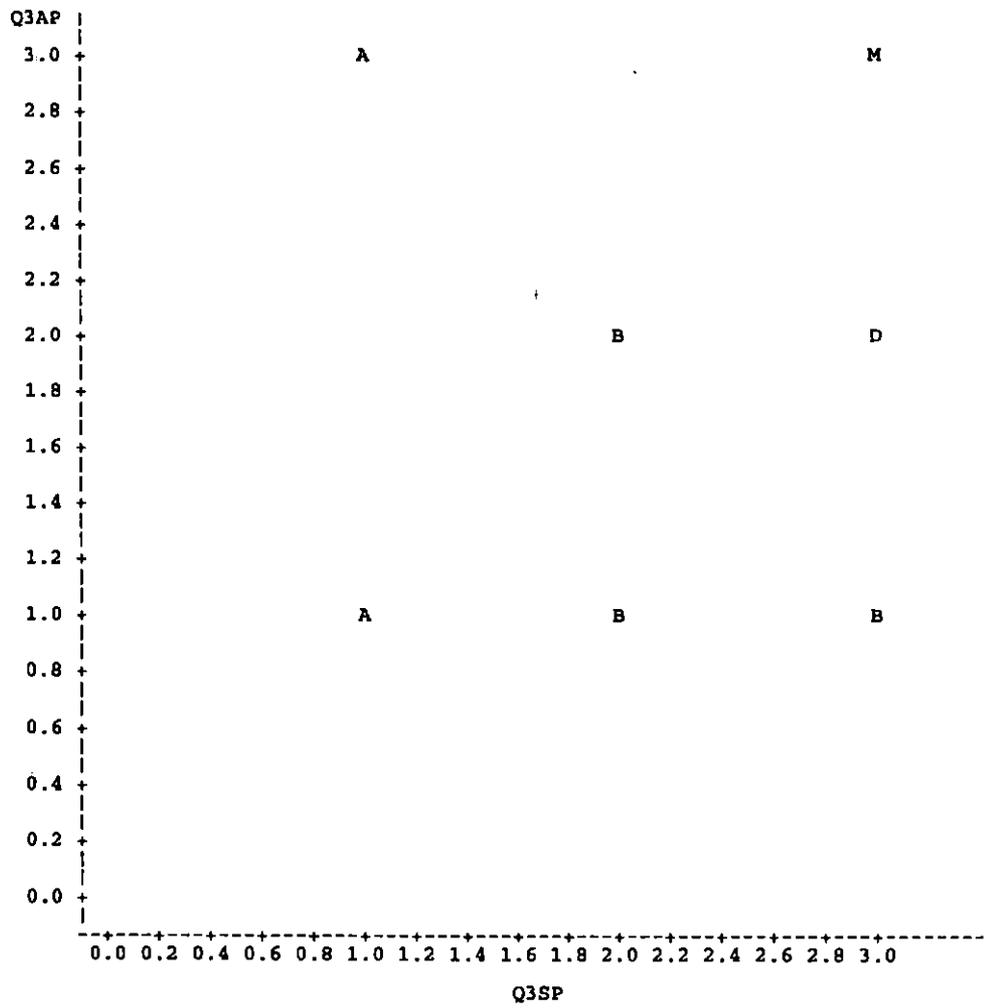
Plot of Q3AP by Q3AS.



Plot of Q3SP by Q3SS.



Plot of Q3AP by Q3SP.



Plot of Q3AS by Q3SS.

