

Copyright
by
Hoonmil Kim
2012

**The Dissertation Committee for Hoonmil Kim Certifies that this is the approved
version of the following dissertation:**

**Writing, Peer Feedback, and Revision: A Comparison of L1 and L2
College Freshmen with Longitudinal Analyses**

Committee:

Carl S. Blyth, Co-Supervisor

Davida H. Charney, Co-Supervisor

Diane L. Schallert

Elaine K. Horwitz

Diana C. Pulido

Christian D. Schunn

**Writing, Peer Feedback, and Revision: A Comparison of L1 and L2
College Freshmen with Longitudinal Analyses**

by

Hoonmil Kim, B.A.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

**The University of Texas at Austin
August, 2012**

Dedication

To my loving family

Acknowledgements

I am indebted to a multitude of individuals in completing this dissertation. This study would not have been possible without their gracious support and sacrifice. Thus, my sincere appreciation and gratitude are due to each and every of them.

My deepest respect and appreciation goes to my Co-Supervisor Dr. Davida Charney. The great classes she taught on composition research inspired this research project. The insightful advice and the invaluable feedback that she provided in every step along the project has shaped this study. From designing the study, recruiting the participants, to running statistical analyses and organizing the chapters of the dissertation, every part of this dissertation is indebted to her invariable support. The meetings and email exchanges with her have always been motivating and informative and have been the major drive that kept me moving forward during this long and challenging journey. I owe my greatest debt of gratitude to her great mentorship, guidance and support.

I would also like to extend heartfelt thanks to my other Co-Supervisor, Dr. Carl Blyth. As my Co-Supervisor, he has provided me with a myriad of valuable suggestions on various ideas and alternative online tools to consider and test for the project. He has also advised me on utilizing technology in organizing and writing the dissertation, and how to coordinate the various feedback of a large committee.

I am humbly indebted to Dr. Diane Schallert, Dr. Elaine Horwitz, and Dr. Diana Pulido, the other members of my dissertation committee. I am grateful for Dr. Schallert's input in the design of 'improve' versus 'non-improve' groups as a means to pursuit research question

three and her advice on utilizing the student survey to complement the findings from longitudinal analysis of feedback comments. She has also motivated me and helped me gain confidence pursuing scholarly works throughout the doctoral program. I owe thanks to Dr. Elaine Horwitz for her invaluable comments that have helped me look at my study from outside of my fixed perspectives. Not only has she been a dedicated member of the committee for the dissertation, but she has also been a truly caring academic advisor who invariably directed me to pertinent courses and helped me focus on academic growth by bringing my attention to potential financial support opportunities. My appreciation also goes to Dr. Diana Pulido for her great Methods class during which the outline for the literature review for the dissertation was developed.

I would like to extend appreciation to my external committee member Dr. Christian Schunn for his indispensable input in constructing survey questionnaires and setting up the analysis scheme for computing the feedback segments.

A special note of thanks goes to two instructor-participants Anthony Arroyo and Mary Hedengren who took extra time and effort in incorporating SWORD into their syllabi. They are enthusiastic and dedicated teacher-researchers who were willing to go an extra mile to try out new technology and methodology in their classes for the potential benefits to their students. I also extend my appreciation to 58 student participants who consented to participate in the study allowing me to review and analyze their data. Without their participation, this study would not have begun.

I owe my loving and unbounded gratitude to my husband Sunggil and my two kids Taehoon and Yejin. As I took the long journey of the doctoral program, they sacrificed their own needs and tolerated the absence of their wife and mother. They supported me

unconditionally each in their own way. Without their understanding and sacrifice, I would not have been able to start, continue, nor complete this task.

I am obliged to my father who has been my role model who gave me the courage to pursue this academic quest. He has given me helpful advice on numerous practical matters including balancing parenting and pursuing a Ph.D, drawing on his own experience.

Loving gratitude also goes to my brother and two sisters who have shown me nothing but unconditional support in everything I did. They have traveled across the U.S. and the Pacific despite their busy schedule to cheer me and offer me a safe and warm haven where I could rest and rejoice. Especially, a special note of thanks goes to my brother Sol for reading this tedious long dissertation and giving me his insightful comments.

Last but not least, I extend my appreciation to my colleague and friend Jayoung Song in the FLE program for graciously helping me fulfill the administrative requirements while completing the dissertation from overseas in Seoul. Without her help, holding an oral defense and distributing the copies of the dissertation to the committee members would have been extremely challenging.

My deepest appreciation for all those above is beyond expression by words.

Writing, Peer Feedback, and Revision: A Comparison of L1 and L2 College Freshmen with Longitudinal Analyses

Hoonmil Kim, Ph.D.

The University of Texas at Austin, 2012

Co-Supervisor: Carl S. Blyth

Co-Supervisor: Davida H. Charney

Abstract: Peer feedback is one of the most popular and widely adopted methods used for writing instruction in both the L1 and L2 classrooms. Previous studies that examined peer feedback suggest different benefits and purposes for the method based on the writers' language group. However, no study has systematically analyzed the peer feedback comments generated by L1 and L2 writers under comparable conditions. While many studies have reported the short-term benefits of peer feedback on writing, little is known in the field about the longitudinal effects of peer feedback on students' writing ability.

This study compares the peer feedback comments of L1 (n=34) and L2 (n=30) college freshman generated in three peer review sessions over a semester using an online peer feedback tool SWORD. Feedback segments (n=4,227) were coded for sixteen feedback features reported to affect the helpfulness of feedback comments. Students' peer feedback profiles were compared between the language groups as well as between the first, second, and third peer review sessions to investigate quantitative and qualitative differences between the language groups and across the feedback sessions. Cases of

students who achieved increase in writing scores over the semester and students with no or negative increase in writing scores were explored in-depth on the feedback they generated, feedback they received, and the revisions they made in order to identify the areas in which they differed.

The results show that contrary to common perceptions, L1 and L2 writers overall generated similar amount and types of feedback comments, with statistical difference found only in the percentage of criticism comments that explicitly stated problems. Students' feedback comments did not change significantly, either in quantity or quality, over time. However, students reported that the feedback they received and provided became more accurate and more helpful over time. Students who achieved an increase in their writing scores behaved differently than those who experienced little or no change in their scores. The improve group made more Type 4 revisions, which is adding/deleting idea chunks, than the non-improve group; the non-improve group received more global criticism feedback than the improve group; little difference was found in the feedback the two groups generated.

Table of Contents

List of Tables	xv
List of Figures	xviii
CHAPTER 1. INTRODUCTION	1
1.1 Background	1
1.1.1 Writing	1
1.1.2 Feedback in Learning and Writing	2
1.1.3 Peer Feedback in L2 vs. L1	4
1.2 Purpose of This Study	5
1.3 Statement of Research Questions	6
1.4 Organization of the Dissertation	7
CHAPTER 2. LITERATURE REVIEW	8
2.1 Revision	8
2.2 Teacher Feedback	12
2.3 Peer Feedback in L1 Studies	15
2.4 Peer Feedback in L2 Studies	19
2.5 Longitudinal Effect of Writing and Revision	22
CHAPTER 3. RESEARCH METHODS	25
3.1 Participants and Settings	25
3.1.1 Setting	25
3.1.2 Participants	25
3.1.2.1 ENG 101 classes	25
3.1.2.2 ENG 101Q classes	26
3.1.2.3 Instructors	27
3.2 Recruiting and Training	27
3.2.1 Instructors	27
3.2.2 Students	27
3.3 Instruments	28

3.3.1 Writing Assignments	28
3.3.2 SWoRD	30
3.3.3 Student Background Survey	32
3.3.4 End-of-Semester Student Survey	33
3.3.5 Instructor Survey.....	33
3.4 Pilot Study.....	34
3.4.1 Need for Pilot Study in ESL class.....	34
3.4.2 Setting and Participants.....	34
3.4.3 Procedures.....	35
3.4.4 Results.....	35
3.5 Data Collection	36
3.5.1 Data from SWoRD: Essays, Feedback Comments, and Back- evaluation.....	36
3.5.1.1 Peer Review Session One	36
3.5.1.2 Peer Review Session Two.....	36
3.5.1.3. Peer Review Session Three.....	37
3.5.1.4 Back Evaluation.....	37
3.5.2. Non-SWoRD Data: Instructors' Grades and Survey Questionnaires	37
3.5.2.1 Instructors' Feedback and Grades.....	37
3.5.2.2 Questionnaires.....	37
3.6 Data Analysis	38
3.6.1 Part 1: Analysis of Feedback Comments	38
3.6.1.1 Coding Scheme for Feedback Comments.....	38
3.6.1.2 Participation Level by Language Group and Gender	42
3.6.1.3 Comparison of Feedback Profiles of L1 and L2 Groups	42
3.6.1.3.1 Missing Data	42
3.6.1.3.2 Statistical Technique: MANOVA.....	43
3.6.1.3.3 Middle-level Coding Scheme	43
3.6.2 Analysis of Effect of Peer Feedback and Revision on Writing over Time	44

3.6.2.1 Analytic Framework	44
3.6.2.2 Sample Case Selection.....	45
3.6.2.3 Themes and Sub-themes	47
3.6.2.4 Theme 1: Types of Revisions Made	49
3.6.2.5 Theme 2: Number of Revisions Made	51
3.6.2.6 Theme 3: Feedback Received	51
3.6.2.7 Theme 4: Feedback Provided.....	53
3.6.2.8 Theme 5: Feedback Incorporation during Revision.....	53
CHAPTER 4. RESULTS and DISCUSSION.....	55
4.1 Research Question 1: Peer Feedback Comments of L1 and L2 Writers.....	55
4.1.1 Students' Level of Participation.....	56
4.1.2 Statistical Test of Differences in Feedback Quantity	58
4.1.3 Gender Effect.....	59
4.1.4 Comparison of Feedback Categories	60
4.1.4.1 High-level Categories	61
4.1.4.2 Detailed Analysis of Criticism Category	61
4.1.4.2.1 Function Subcategory: Problem, Solution, Prob+Sol ..	61
4.1.4.2.2 Scope Subcategory.....	62
4.1.4.3 Specificity	63
4.1.5 Comparison of Textual Features of Feedback	64
4.1.5.1 Use of Modal Verbs	64
4.1.5.2 Use of Mitigation Language	70
4.2 Research Question 2: Longitudinal Changes in Comments	72
4.2.1 Quantitative Developments.....	73
4.2.2 Changes in Feedback Categories	74
4.2.2.1 High-level Categories	75
4.2.2.2 In-depth Analyses of Criticism Category	77
4.2.2.2.1 Criticism by Function: Problem, Solution, Problem+Solution	77
4.2.2.2.2 Criticism by Scope: Macro/Meaning-level vs. Micro/Surface-level	79

4.2.2.3 Specificity: General vs. Specific Feedback	80
4.2.3 End of Semester Student Survey.....	80
4.2.3.1 Changes in Attitude and Perception.....	81
4.2.3.2 Perceived Changes in Incorporation of Feedback	82
4.2.3.3 Perceived Changes in Feedback Skills	85
4.2.3.4 Perceptions of Importance of Grading Peer Feedback ...	86
4.3 Research Question 3: Improve vs. Non-Improve Groups.....	87
4.3.1 Revision	90
4.3.1.1 Revision Repertoire	90
4.3.1.2 Amount of Revision.....	92
4.3.1.3 Use of Type 4 Revision: Idea Addition/Deletion	94
4.3.2. Feedback Incorporation	97
4.3.2.1 Feedback Received	97
4.3.2.1.1 Criticism and Global Criticism Comments Received ..	98
4.3.2.2. Feedback Incorporation	99
4.3.3 Feedback Provided.....	103
4.3.4 Longitudinal Effect of Feedback and Revision on Writing	106
CHAPTER 5. CONCLUSION.....	111
5.1 Summary of the Findings.....	111
5.1.1 Research Question One.....	111
5.1.2 Research Question Two	114
5.1.3 Research Question Three	116
5.2 Implications.....	121
5.2.1 Theoretical Implications	121
5.2.2 Methodological Implications	122
5.2.3 Pedagogical Implications	124
5.3 Limitations	126
5.4 Future Research	127
Appendices.....	129
Appendix A. Student Background Survey.....	129

Appendix B. Students' Post Semester Survey	133
Appendix C. Instructor Survey	142
Appendix D. Slides of Examples of Good and Bad Feedback	148
Appendix E. Example of Count and Categorization of Revision Strategies	152
References.....	154
Vita.....	165

List of Tables

Table 3.1 Demographic and Background Information of Participants	26
Table 3.2 Definition and Examples of Coding Categories	39
Table 3.3 Middle-level Categories and Their Sub-variables	44
Table 3.4 Background and Writing Scores of Sample Cases	46
Table 3.5 Themes and Sub-themes	48
Table 3.6 Typology of Revision Strategies.....	50
Table 3.7 Example of Feedback Received, its Categorization and Counting.....	51
Table 4.1.1 Students' Participation Level in Peer Review Sessions.....	56
Table 4.1.2 Comparison of Number of Reviews, Idea Units, and Words	58
Table 4.1.3 Gender Effect on Number of Reviews, Idea Units and Words.....	60
Table 4.1.4 Feedback Categories Used for Coding	60
Table 4.1.5 Distribution among Praise, Criticism, and Reflection Comments.....	61
Table 4.1.6 Comparison of Sub-categories of Criticism.....	62
Table 4.1.7 Comparison on Meaning-level vs. Surface-level Feedback	63
Table 4.1.8 Distribution of General vs. Specific Feedback	64
Table 4.1.9 Comparison of Use of Modals in Solution Comments	64
Table 4.1.10 Examples of L2 students' Use of <i>Should</i> and <i>Can</i>	68
Table 4.1.11 Examples of L1 Students' Feedback with Mitigation	70
Table 4.1.12 Examples of L2 Students' Criticism Feedback without Mitigation	72
Table 4.2.1 Mean Percentage Changes in Praise, Criticism, Reflection over Time.....	75
Table 4.2.3 Changes in Problem, Solution, Problem+Solution over Semester	79
Table 4.2.4 Changes in Meaning-level versus Surface-level Feedback over Semester	80

Table 4.2.5 Changes in General vs. Specific Feedback	80
Table 4.2.6 Mean Responses on a Student Survey on Behavior and Attitude towards Peer Feedback on a Six-Point Likert Scale with 1 = Strongly Disagree and 6 = Strongly Agree	82
Table 4.2.7 Percentage of Reported Feedback Incorporation by Group and by Session	83
Table 4.2.8 Reasons for Change in Feedback Incorporation over Time	84
Table 4.2.9 Students' Perceived Improvement in Peer Feedback Performance	85
Table 4.2.10 If Peer Review/Feedback is Graded.....	87
Table 4.3.1 Background and Writing Scores of Sample Cases	87
Table 4.3.2 Themes and Sub-themes	88
Table 4.3.3 Summary of Topics and Findings	90
Table 4.3.4 Comparison of Revision Repertoire	91
Table 4.3.5 Comparison of Number of Total Revisions by Group.....	92
Table 4.3.6 Changes in Amount of Revision across Sessions (L1)	93
Table 4.3.7 Changes in Amount of Revision across Sessions (L2)	94
Table 4.3.8 Breakdown of Revisions by Type (L1).....	94
Table 4.3.9 Breakdown of Revisions by Type (L2).....	95
Table 4.3.10 Revision with Type 4 Strategy.....	96
Table 4.3.11 Number of Criticism and Global Criticism Comments Received by Group	98
Table 4.3.12 Percentage of Feedback Incorporation by Group	99
Table 4.3.13 Comparison of Feedback by Improve and Non-improve Groups (L1)	104
Table 4.3.14 Comparison of Feedback by Improve and Non-improve Groups (L2)	104
Table 4.3.15 Virtual Scores on Six Drafts of Essays	106

Table 4.3.16 Scores and Revisions by Type across Sessions (L1)	109
Table 4.3.17 Scores and Revisions by Type across Sessions (L2)	109

List of Figures

Figure 4.1 Changes in Number of Feedback Ideas and Words across Sessions	73
Figure 4.2 Changes in Problem, Solution, Problem+Solution over Semester	79
Figure 4.3 Changes in Virtual Scores on Six Drafts of Essays over Time	107

CHAPTER 1. INTRODUCTION

This study investigated undergraduate students' peer feedback and revision activities in relation to development of their writing competency over one semester. Specifically, the feedback comments of L1 and L2 students were compared in terms of quantity and quality to understand the efficacy of their feedback comments as a viable alternative source to teacher feedback. The changes in the feedback comments were tracked over the semester and compared across sessions and between the groups to identify the effect of repeated engagement in peer feedback activity on students. The study also explored components of peer feedback and revision activities that were associated with improving students' writing proficiency over the semester.

1.1 BACKGROUND

1.1.1 Writing

One's ability to compose an extended text is the single best predictor of success in course work during the freshman year (Geiser, 2001). Gains in informative and analytical writing ability are taken as a good indicator of the value added by higher education (Benjamin & Chun, 2003). Moreover, effective writing skills are critical for both entering the workforce and as a means for advancement (National Commission on Writing, 2004).

Although educators agree on the need to equip students with adequate writing ability through education, many students in the United States still appear to lack the writing skills needed for college (National Assessment of Educational Progress, 2007). In fact, some college graduates are also unprepared for the advanced writing tasks required in the workplace (National Commission on Writing, 2004).

Although large-scale statistics on the level of writing skills of English as Second Language (ESL) learners are unavailable, partially due to the wide variability of skills demonstrated by L2 learners at different stages of their language development, studies that

compare the written texts of L1 and L2 writers commonly report lower holistic scores on L2 writers' texts (Campbell, 1978a, 1987b, 1990; Carlson, 1988; Connor, 1984; Hafernik, 1990; Park, 1988; Reid, 1988; Santiago, 1970; Xu, 1990). Additionally, L2 writers' texts are found to display distinct features that are associated with less developed writing such as less fluency (Benson et al., 1992; Cummings, 1990; Hall, 1990; Lin, 1989; Silva, 1991, Reid, 1988; Yu & Atkinson, 1988) and more errors (Benson, 1980; Frodesen, 1991; Silva, 1990; Stalker & Stalker, 1988). These reports suggest that the writing competency of ESL/EFL learners is likely to fall short of the standards of NAEP, even more so than is the case of L1 students. Yet, in this highly globalized world, many ESL/EFL learners are enrolled in academic institutions where the medium of instruction is English, taking courses that require written skills comparable to that of native English speakers. Moreover, most of these students seek to enter job markets that require excellent English writing skills after graduation. Thus, writing development for L2 learners is an important issue in the ESL/EFL academic community.

Scholars have pointed to insufficient practice as one of the core problems causing students' lack of writing skills (Cho & Schunn, 2007; Kellogg & Whiteford, 2009). The major impediment to engaging students in enough writing tasks is the extensive time and effort involved in grading papers and providing feedback (Graham & Perin, 2007; Kaufman & Schunn, 2007; Kellogg & Whiteford, 2009). As a way to alleviate the problem, some scholars have suggested employing alternative sources and methods for providing feedback to students' writing in addition to the traditional hand-written feedback from teachers.

1.1.2 Feedback in Learning and Writing

Feedback is information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding (Hattie & Timperley, 2007). It is also defined as "information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies" (Winne & Butler, 1994: p. 5740).

Feedback is one of most powerful instructional methods to aid students' learning. Hattie (1999) conducted a synthesis of over 500 meta-analyses involving 180,000 studies to identify the effect of feedback on learning in comparison with other influences. Analysis of 196 studies involving feedback showed that the effect size of feedback on learning is twice as large as the standard effect size of schooling on learning (0.79 vs. 0.40). However, there was wide variability in the effect size based on the types of feedback employed. Thus, although feedback is effective in bringing growth in students in general, some forms of feedback are not as effective as others.

In writing instruction, although feedback has long been provided in composition classrooms, it was with the rise of the process-oriented writing curriculum in the 1970s that interim as well as final feedback and feedback from varied agents became prominent in American schooling. As the instructional focus shifted from the product to the process of writing, more and more writing instruction involved students in producing multiple drafts guided by teacher and peer feedback. Reviewing the history of teaching composition at postsecondary level, Haswell (2005) commented that peer review and feedback were embraced with enthusiasm by scholars interested in college writing instruction and are now very common in college writing classrooms. The wide usage of feedback attracted heightened academic attention, resulting in a proliferation of studies systematically investigating the effect of feedback on writing.

For example, Hillocks (1984) performed a meta-analysis of over 500 empirical studies intervening on writing to unveil what works in teaching composition. He examined the mode (method) of instruction and the focus (content) of instruction as the two dimensions for analyses. He found that in terms of mode, engaging students with each other and activities conducive to high levels of peer interaction had higher effect size (0.75) than other modes. As for the focus of instruction, involving students to use inquiry and criterion scales was found most effective with effect sizes of 0.56 and 0.36. Hillocks' findings lent support to peer feedback, where it engages students in active learning and encourages discussion of rubrics. Beason (1993) also noted that "feedback and revision are valuable pedagogical tools. . .the

research typically indicates that high school and college students improve their drafts upon receiving feedback” (p. 396). Others reported that teacher feedback is perceived as helpful by students (Radeki & Swales, 1988) and that peer feedback as mostly valid (Rollinson, 1998; Caulk, 1994).

The factors that make feedback effective have not been settled. Some studies found teacher feedback to be vague (Sommers, 1982) and arbitrary (Zamels, 1982), and student feedback to be inaccurate (Chou, 1999; Leki, 1990; Mendonca & Johnson, 1994) and not conducive to revision (Connor & Asenavage, 1994; Chou, 1999). Students have been criticized for the lack of knowledge and skills necessary to provide peer feedback (Flynn, 1982; Wiener, 1986). These mixed results are partially due to variability in study setting, task type, and task procedures.

More recently, a series of studies carefully controlled for the variability in the format and the procedures of peer feedback sessions by employing a web-based peer feedback tool named SWoRD (Scaffolded Writing and Rewriting across Discipline; Cho & Schunn, 2007; Cho et al., 2006; Patchan et al., 2009). SWoRD was developed specifically to aid students’ peer review activity based on the best practices found to be effective in improving students’ peer review activity. Also, using SWoRD in peer review sessions allows control over the condition and environment of peer review sessions by use of standard procedures such as use of common rubrics. Feedback studies using SWoRD has reported relatively consistent findings on students’ and teachers’ feedback comments, allowing more accurate comparison of the results across studies. Peer feedback gained stronger support as pedagogy for writing with its reliability, validity, and effectiveness consistently reported to match or exceed that of teacher feedback.

1.1.3 Peer Feedback in L2 vs. L1

Backed by social learning theory, process pedagogy was soon introduced in L2/ESL writing environment. However, many teachers in the L2 community continued to believe that students are incapable of evaluating peers because of their lack of language ability, skill, and experience (Saito & Fujita, 2004) and that L2 learners may prefer teacher feedback to peer

feedback (Zhang, 1995). However, studies that investigated the effect of training on peer review (Stanley, 1992; Berg, 1999; Min, 2005, 2006) showed that after training the feedback comments of peers became more specific and relevant, got incorporated more during revision, and addressed meaning-level features. These studies point to the possibility that L2 learners can be trained to produce peer feedback comments that are accurate/valid and helpful for improving writing, and to the possibility of using peer feedback in lieu of teacher feedback that is in short supply in the educational settings.

No studies to date have compared L1 and L2 students using the same feedback protocols on comparable writing tasks over the course of a term. It will be of vast interest to the L2 community to find out how the findings in the L1 literature can relate to and interpreted in the L2 context. The unique features of SWORD suggest that if L2 students were to use SWORD for their peer review sessions, their feedback may resemble that produced by L1 students who used SWORD.

1.2 PURPOSE OF THIS STUDY

The purpose of this study was three-fold. First, this study re-examined how peer review can be employed in L2 writing classrooms. Previous studies in L1 suggest (Cho & Schunn, 2007; Patchan, Charney & Schunn, 2009) that peer feedback can be an alternate source for teacher feedback because its quantity and quality are similar to teacher feedback. On the other hand, studies in L2 (Miao et al., 2006) propose using peer feedback as a supplement to teacher feedback as each form of feedback offers distinct benefits. Against this backdrop, this study compared the quantity and the quality of feedback comments generated by L1 and L2 students under similar conditions using SWORD. The findings from this analysis not only provide guidance as to whether the pedagogy of peer feedback has a different role in the two language groups but also offer insights on how the findings reported in L1 SWORD studies can be related to and applied to an L2 setting. The results may have practical implications for designing curricula, syllabi, or lesson plans for peer review and feedback activities that cater to the specific needs of each language group. Comparing the two groups of learners is crucial in today's diverse educational settings where many

classrooms across the globe accommodate student population composed of both L1 and L2 learners.

Second, the study investigated the longitudinal aspects of peer review activities. In natural classroom settings, peer review is often employed multiple times throughout the semester. Nevertheless, few studies have explored the developments that take place over time. This study examined how students' engagement in peer review activities changed over time and how the actual feedback comments developed across sessions. Findings on the effects of repeated peer feedback sessions provide useful information for designing curricula and programs that involve peer feedback.

Third, this study aimed to unveil how peer feedback and revision affect development of students' writing competency in the long term. So far, studies that looked into the relationship between writing development and feedback/revision have only focused on short-term effects of peer feedback on writing. Studies have shown that peer feedback generally leads to improvement in writing quality from the first to the next draft. However, it has seldom been reported in the literature whether this improvement between the drafts leads to improvement in writers' general writing ability that will carry on beyond the current drafts. Ultimately, students must be able to write well even when there are no teachers or peers to provide helpful feedback.

1.3 STATEMENT OF RESEARCH QUESTIONS

Based on the review of the studies conducted to date on the topic of peer feedback and writing, the following three research questions were formulated to guide the current study.

Research Questions:

1. How do peer feedback comments generated by L1 writers compare to those of L2 writers when produced under comparable conditions using the web-based peer review commenting tool SWORD?
2. How do peer feedback comments of L1 and L2 students change/develop over a semester?

3. How do the feedback and revision activities of students whose writing improved over the semester compare to those of students whose writing did not improve? How do L1 and L2 students compare within these two achievement groups?

1.4 ORGANIZATION OF THE DISSERTATION

In this chapter, the background and the purpose of this study, and the research questions were presented. The following chapter reviews literature dealing with the topics of teacher and peer feedback, revision, and writing. Chapter Three presents the methodological procedures and concerns in designing this study and analyzing the data. In Chapter Four, the findings from the data analysis are presented and discussed. Finally, Chapter Five provides a summary of the findings, the implications of this study from theoretical, methodological, and pedagogical perspectives, the limitations of this study, and suggestions for future study.

CHAPTER 2. LITERATURE REVIEW

2.1 REVISION

Revision is defined as any change(s) to any part of existing text at any point during the writing process (Flower, Hayes, Carey, Schriver & Stratman, 1986). Revision is commonly regarded as a central and significant part of writing (Lowenthal, 1980; Murray, 1978a; Scardamalia & Bereiter, 1986). Revision is important because, under certain circumstances, it brings improvement to the quality of the final written work (Ash, 1983; Bamberg, 1978; Bracewell, Scardamalia & Bereiter, 1978; Bridwell, 1980). Scholars (Flower et al., 1986) note that revision is the profession's method of choice as teachers assert that the practice of multiple drafts via revision is the key to good writing.

Despite the reported virtue of revision on improving writing quality, studies suggest that students are not always successful in making good use of this process. Some researchers investigated whether revisions are made where they were needed and whether it improved the quality of writing. Studies report that by and large students don't revise (NAEP, 1977), and if they do, their texts may get worse from it (Perl, 1979; Beach, 1976). Faigley and Witte (1981) compared writers' choices and decisions for revision on the same draft of writing between expert adult writers and inexperienced adult writers and found little correspondence between the two groups. Scardamalia and Bereiter (1983) compared the revision choices of fourth, sixth, and eighth graders against semi-professional writers and found that the revision tactics of fourth graders matched the least, sixth graders' moderate, and the eighth graders' the most with the semi-professional writers. Hayes et al. (1987) found that college freshman tend to focus their revision activities on problems at or below the sentence level while more advanced writers attend both to local and global problems. The studies above suggest that revision is a difficult task especially for novice and student writers.

There are a number of reasons that account for the failure to revise successfully. According to Hayes, Flower, Schriver, Stratman and Carey's (1987) model of revision,

revision consists of two major processes: evaluation and strategy selection. Central to this model is the evaluation function (Hayes, 1996), a process that is responsible for the detection and diagnosis of text problems. Using the model, Flower et al. (1986) identified three points in the process of revision that might be the source of trouble for writers: detecting that a problem exists; building a diagnostic representation of the problem; and selecting a strategy for revision. Each of these points may be a barrier to students that bars them from conducting revision successfully.

The first step in the process of revision is detection. Detecting problems during revision is an act of sensing the dissonance between the writers' intention and text (Flower et al., 1986). Without detection, subsequent steps in the revision process cannot take place. Detection is a necessary and potentially difficult process. Rubin (1984) reports that even the "A" students missed 75% of the problems their instructor identified in a study that examined students' self-evaluation. Hayes et al. (1987) found that even experts identified only 66% of the planted problems during revision when shown by one sentence at a time, which is yet higher than the 42% detection rate of novice writers.

Two things account for students' failure to detect problems in the text: creating inaccurate representation of the text and applying narrow task definition of revision and low standards for good writing. To understand whether students find it more difficult to identify problems in their own writing than others', Bartlett (1981) compared students' performance on detecting missing subjects or predicates and detecting faulty referents when they occurred in other students' texts versus when they were in their own texts. She found that students' performance on detecting the ambiguous references dropped from 50% to 10% when they were in their own texts while the performance on the rule-governed errors stayed the same. This study suggests that problem detection is affected by writers' ability to provide an accurate representation of the text, separate from their own internal representation of meaning.

In addition, the goals and criteria the writer brings to the task of revision affects the types and range of problems that are detected. Students in Sommers' study (1980) who were

asked to make revisions, only deleted unnecessary words in the text. The lawyers in Bond's study (1980) who were asked to revise a complex legal document for lay readers made only minimal changes to the text. These revisers detected only the events that violated their own internal standards of good writing that they brought to the revision task.

While detection is a necessary step for revision, detecting a problem merely means that the reviser acknowledges that a problem(s) exists in the text. After successfully detecting problems in the text, students may still face difficulty in the subsequent step of revision, or diagnosing the problem. Flower et al. (1986) claim that the representation of problems writers create exists along a continuum with simple detects such as 'this just doesn't sound right' on the one end and well defined problems such as 'faulty parallelism' on the other. The degree of specificity and sophistication of problem representation determines the difference between "detect" and "diagnosis." In order for the reviser to act upon the problem, enough information has to be supplied through diagnosis of problems that go beyond simple detection. For example, Scardamalia and Bereiter (1983) found that children writers in their study were as good as adult writers in choosing general problem definitions but were not able to define the problems more explicitly nor select a means of dealing with it. Flower et al. (1986) explains that diagnosis differs from detection in that it points the way to a solution and asserts that novice and expert writers differ in the capacity to employ detection and diagnosis for different reasons and at different times.

Finally, the third step in the revision process is selecting a strategy. Flower et al. (1986) presents two major paths of action as strategy: the Detect/Rewrite strategy and the Diagnosis/Revise strategy. A mere detection offers limited choices and commits the writer to the Rewrite strategy while diagnosis allows the writer to access a variety of strategies tied to the problem (s)he defines. While rewriting may be a strategy necessary in some cases, it may lead to a relatively blind leap into ill-defined problems as it fails to recognize that a definable problem may exist. Based on their studies, Flower et al. (1986) claim that experts rely heavily on diagnosis to help them recognize and plan around problems in order to carry out global and whole-text revisions. From these findings, it can be argued that students need to be

guided to take a Diagnose/Revise strategy as they begin making global and whole-text revisions (Wallace & Hayes, 1991).

All these findings point to the need to help inexperienced writers in revision. Some intervention research suggests that feedback may enhance the link between revision and writing quality. Buxton (1959) found that a group of college freshman who received feedback before revising outscored the other group who did not receive feedback on writing scores. Ferris (1997) examined 1,600 marginal end comments written on 110 drafts of papers to investigate characteristics of teacher feedback that influences student revision and whether revisions made as result of teacher commentary lead to substantive changes. Results showed that in general, longer comments and text-specific comments were associated with major changes than shorter, general comments. When changes were made, they tended to improve students' papers. Other studies also report that teacher or peer feedback can enhance revision for writers in the primary grades through high school, especially if the feedback is focused (Gere & Stevens, 1985; Hillocks, 1982; Sperling & Freedman, 1987).

More recently, researchers have come up with various ways to help students overcome difficulties they may face in each step of the revision process. As a way of helping students identify problems in the text better during the detection process, Schriver (1992) suggested training students to anticipate readers' needs. In a study that investigated the effect of reader-protocol teaching method, it was found that the method significantly increased students' ability to diagnose readers' problems, characterize problems from the readers' perspective, and attend to global text problems. Wallace and Hayes (1991) posited that the student writers may have a narrow definition of revision task that leads to making only local changes below the sentence level and used an eight minute instructional prompt to encourage students to revise globally. The study found that the prompt stimulated both an increase in the amount of global revision and an improvement in text quality.

All in all, the literature on revision suggests that revision is a challenging task for student-writers and that the use of training and feedback in writing classrooms is needed to enhance students' revision performance. While the effect of teacher feedback on student

revision has been investigated by several studies (Ferris, 1997; Straub, 2000), the effect of peer feedback on student revision has been scarcely reported in the field (cf. Nelson & Schunn, 2008). With this gap in the body of knowledge, the present study will investigate the types of revisions made as a result of peer feedback in L1 and L2 writing classrooms and its effect on the writing quality.

2.2 TEACHER FEEDBACK

Teacher feedback has been one of the most popular and commonly used methods of instruction for writing throughout the history of education. Over the decades, a myriad of studies were conducted to explore how teacher feedback affects students' writing and under what conditions teacher feedback is most effective in bringing growth in students' writing. Teacher feedback sets the model by which students' peer feedback is performed, investigated and interpreted. Earlier studies examining teacher feedback show mixed results regarding the benefits of teacher comments, but more recent studies focus on specific dimensions of teacher commentary and offer ways to improve the efficacy of teacher feedback in classrooms.

Studies show that not all teacher-generated feedback is helpful and valid. In one of the earlier studies that examined teacher commentary, Sommers (1982) analyzed written feedback of thirty-five teachers and found that most comments were not text-specific but a series of "vague directives" that could be "interchanged, rubber-stamped, from text to text" (p. 111). Similarly, Sperling and Freedman (1987) reported that one student consistently misinterpreted her teacher's marginal comments because the student did not share the teacher's knowledge and values regarding writing and revision. In the L2 domain, Zamel (1982) examined fifteen teachers' feedback on 105 students' essay and reported that most teachers focused on surface-level errors, made arbitrary corrections, gave contradictory comments and rarely suggested specific solutions or strategies for revision. In another L2 study, Lee (2011) investigated the feedback practice of 26 teachers in Hong Kong and found that these teachers in EFL settings predominantly focused on error correction.

However, students believe that teacher comments are valuable. In response to a survey conducted by Lynch and Klemans (1978) targeting 154 L1 college students, 92%

answered that teacher comments were helpful. Radeki and Swales (1988) asked L2 students about the usefulness of different types of teacher commentary; most reported that their teacher's comments were either helpful or somewhat helpful. Straub (2000) surveyed L1 students on the effects of different types of teacher feedback on revision. Students preferred comments that provided advice, included explanations, and employed open-ended questions. Rather than focusing simply on student perceptions, Ferris (1997) analyzed the revisions students make as a result of teacher feedback and found that marginal comments, requests for clarification, and comments on grammatical errors led to the most effective revisions for L2 students.

More recently, researchers have conducted in-depth analyses of teacher commentaries on specific dimensions and suggested ways to improve the effectiveness of teacher feedback. For example, Summer Smith (1997) analyzed the topics and sequence of 208 end comments of ten teachers of first-year composition and rhetoric courses. She found that teachers' end comments show the stability of a genre with specific features in content and sequence that make end comments recognizable. While these standardized features may make it easier for the teachers to write the end comments, the researcher claims that it may also reduce the educational effectiveness of the comment. Summer Smith posits that if students do not read comments carefully, it may be because the comments take highly standardized forms and calls for teachers to heighten awareness of the constraints of generic conventions and the danger they pose to the effectiveness of the feedback.

In another study, Treglia (2009) analyzed the phrasing of teachers' feedback to investigate whether students misunderstand mitigated commentary. She collected students' drafts from two mixed classes consisting of 22 L1 and 21 L2 students. L1 and L2 students had no problem understanding the intent of mitigated commentary and successfully addressed the issues raised by it. She concluded that the type of comment, linguistic form, and hedging used by the teacher were not the determining factors for poor revisions. Instead, students failed to address feedback because of the nature of the problem being addressed. That is, students had difficulty addressing comments that presented challenging analytical tasks such

as re-evaluating the logic of a passage, explaining an idea, or connecting ideas. The study suggests that researchers and teachers need to focus more on providing specific guidance on helping students improve their ability to make these challenging revisions than simply attending to the features of feedback comments.

These studies indicate that the question of whether teacher feedback is helpful or not to students cannot be collectively answered for all teachers' end comments despite the common characteristics they share as a specific genre of writing. Depending on how much the comments are customized to the specifics of each writing task, the types of features they address, and how many solutions and suggestions they contain, the degree of helpfulness will vary.

Studies also indicate that teacher comments can differ depending on expertise in subject area and expertise in writing instruction. Patchan, Charney and Schunn (2009) conducted an in-depth analysis of students' comments against the comments of a writing instructor and a content instructor. Over 1400 comment segments generated by undergraduates, a writing instructor and a content instructor were coded for the presence of 29 different feedback features. Some of the differences among the three participant groups include students using praise twice as often as the instructors; the content instructor explicitly identifying more problems than the students, who did so more than the writing instructor; both instructors identifying issues related to the content of the paper more often than the students. Lastly, the content instructor most frequently provided solutions regarding the content, the writing instructor most frequently provided solutions regarding high prose, and students most frequently provided solutions on low prose. The data not only showed that the content and the writing instructors vary in how they provide feedback but more importantly, that the student group overall fell in between the two instructors in most of measures, indicating that students' comments were fairly similar to the instructors' comments. This is presumably because students have more content knowledge than the writing instructor and have been exposed to writing feedback more frequently or recently than content instructor.

The authors concluded that under certain conditions, students provide comments that are similar in both quantity and quality to those of instructors.

The fact that teacher feedback can vary so much in type and effectiveness raises the attractiveness of finding a mechanism to provide the most effective forms of feedback to students in classrooms. Given the wide variation in teacher feedback, an even greater variability is likely to exist in the feedback comments produced by peers, which partially explains why aggregated feedback of at least four peers are found to be reliable and valid (Cho et al., 2006). As such, it may be difficult to generalize helpfulness, or lack thereof, of peer-generated feedback in absolute terms. Instead, studies that investigate the specific features of peer feedback that are associated with revision and writing quality are called for as these studies will be suggestive of ways to improve the efficacy and helpfulness of peer feedback comments.

2.3 PEER FEEDBACK IN L1 STUDIES

Peer review/response is defined as “the use of learners as sources of information, and interactants for each other in such a way that learners assume roles and responsibilities normally taken on by a formally trained teacher, tutor, or editor in commenting on and critiquing each other’s drafts in both written and oral formats in the process of writing” (Liu & Hansen, 2002:1). Traditionally, peer feedback has been widely used in composition classrooms in the U.S. with various benefits claimed to be associated with it. Theoretically, peer review is supported by several frameworks, including process writing, collaborative learning theory, and Vygotsky’s social learning. Peer response is considered a necessary component in the process writing approach that emerged in the 1970s (Elbow 1973; Emig, 1971). It is also supported by the collaborative learning theory, which holds that learning is a socially constructed activity that takes place through communication with peers (Bruffee, 1984). Vygotsky’s Zone of Proximal Development theory (1978), which maintains that cognitive development of individuals results from social interactions in which individuals extend their current competence through the guidance of more experienced individual vis-à-vis scaffolding also supports peer review.

More recently, a series of studies empirically investigated the efficacy and benefits of peer feedback. First, researchers examined the efficacy of feedback comments generated by peers by testing the reliability and validity. One of the major concerns of teachers and students on using peer feedback as a viable source of feedback has been the possibility of low reliability and validity of evaluation made by students (Cho et al., 2006). Cho, Schunn and Wilson (2006) examined grades given by 708 students across 16 different courses from four universities. The students were given guidance on peer assessment, used rubrics, and were provided with clear incentives to take the assessment seriously. The mean ratings on students' writing generated by peers were compared against those given by instructors to test reliability and validity from the instructors' point of view. The analyses suggest that the aggregate ratings of at least four peers on a piece of writing are both highly reliable and valid as instructor ratings. In another study, Cho, Schunn and Charney (2006) examined the characteristics and helpfulness of peer generated feedback in comparison to that of instructor feedback. The researchers compared the feedback comments of undergraduate and graduate students with a subject matter expert and found that the subject expert's comments were significantly longer, had more specific suggestions, and contained less praise comments than the students' comments. Although the feedback comments produced by different reviewer groups showed different features, no significant difference was found between the peer and instructor feedback in terms of helpfulness perceived by writers. Together, these studies suggest that although the feedback comments generated under appropriate conditions by peers may contain relatively more praise, are shorter, and have less specific solutions compared to teacher feedback, they are as valid and reliable as teacher feedback and are perceived as equally helpful in revision by student-writers.

While these studies attended to the different characteristics of feedback comments generated by various reviewers, Nelson and Schunn (2008) investigated how different types of peer feedback affect students' writing performance. To address this question, 1,073 feedback segments were coded for eight features: summary, problem, solution, localization, explanation, scope, praise, and mitigating language. The relationship between the feedback

features and implementation of feedback during revision was analyzed. The results suggest that solution was the only feature that had statistically significant relationship with implementation. A writer was 10% more likely to make changes when a feedback comment included a solution than when it did not. In addition, understanding was found to be a significant mediator of implementation, where solution, summary, and the location of the problem were associated with increased understanding. The study points to the possibility of increasing effectiveness of peer-generated feedback by guiding students to generate feedback comments that includes features that promote implementation.

Patchan and Schunn (2011) investigated the effects of receiving different types of feedback on revision and on the quality of final drafts. They studied the quality of initial draft, feedback comments provided, revisions made and the quality of final draft of 211 undergraduate students in a science course. The students wrote the initial drafts and received feedback from either a TA (Teaching Assistant) of the course or four other peers. Analyses of the data show that the comments produced by TAs were generally shorter, contained less prose-related comments, and had similar amount of content-related feedback compared to the mean of feedback given by peers. Students who received feedback from peers received significantly more comments when combining the feedback from four peers and made significantly more changes to the text during revision than the students who received feedback from a single TA. However, when the types of changes made were compared, students who received peer feedback made more low prose changes but the similar number of high prose and content changes than the students who received TA feedback, indicating that TAs' feedback is more economical than peer feedback.

Taken together, these studies indicate that when peer feedback is generated under certain conditions, feedback comments from multiple peers can be as reliable, valid and helpful as teacher feedback. Although peer generated feedback comments may not contain the same features as the feedback by teachers when compared one to one, when used collectively, peer feedback seems to qualify as an alternate source to teacher feedback, that

can increase the opportunity for students to write more, which has been pointed out as one of the solutions for students' poor writing performance.

While a majority of studies on peer feedback examined the effect of receiving peer feedback, a number of studies explored the benefits students gain from reviewing and evaluating writing of peers. Bruffee (1978) examined the writing of tutors over a semester and found that both the tutors and the tutees they worked with improved significantly in their writing over the semester. Graner (1987) compared two writing classes in which students only reviewed others' papers in one class while students in the other class both reviewed and received comments from peers. He found that the students who reviewed papers but did not receive any feedback improved at the same rate as students who both received and gave feedback. Cho and Cho (2010) investigated the relationship between providing feedback on peers' writing and improvement in the reviewers' own writing. The qualities of initial and final drafts of papers written by 72 undergraduate students along with 3,889 segments of peer feedback comments were analyzed. Multiple regression analyses found that the reviewers' comments significantly influenced their own revisions. When the student reviewers commented more on the strengths of macro-meaning and the weaknesses of micro-meaning features, the revision qualities of their own drafts tended to improve. In addition, reviewers' initial writing skills and the quality of reviewed peer drafts were found to influence the types of comments given.

All in all, studies on peer feedback suggest that students have much to gain from peer feedback activities from both providing and receiving feedback when peer feedback is conducted under certain conditions. These conditions include training students on the task schema of revision to include revisions on macro and global features of texts; providing students with well defined rubrics for evaluation of paper; and providing incentives to students to take peer feedback and revision activities seriously. While doubts still linger on whether peer feedback is a reliable and valid source of feedback in writing classrooms, as more research is conducted on the topic, peer review and feedback is gaining a firmer ground as a valid alternate source to teacher feedback, at least in the L1 domain.

2.4 PEER FEEDBACK IN L2 STUDIES

Various claims have been made about the cognitive and social benefits of engaging students in peer review and feedback in the L2 community. Peer review has been reported to help both college (de Guerrero & Villamil, 1994; Medonca & Johnson, 1994; Villamil & de Guerrero, 1996) and secondary (Peterson, 2003; Tsui & Ng, 2000) students gain more insight into their writing and revision processes, bring a sense of ownership of the text (Tsui & Ng, 2000), foster more positive attitudes toward writing (Min, 2005), increase audience awareness (Mendonca & Johnson, 1994; Mittan, 1989; Tsui & Ng, 2000), and provide students with an opportunity to practice listening and speaking skills (Lockhart & Ng., 1995; Mendonca & Johnson, 1994).

Despite the cognitive, affective and social benefits of peer feedback, studies that investigated the efficacy of peer feedback report mixed findings in the L2 domain. For example, Nelson and Murphy (1993) identified problems concerning the depth, accuracy and credibility of peer feedback. In a similar vein, a number of studies noted that feedback comments generated by L2 learners tend to give rubber stamp advice to peers (Chou, 1998; Leki, 1990; Mendonca and Johnson, 1994; Lockhart and Ng., 1993).

On the other hand, a study by Paulus (1999) analyzed the types of feedback peers and teacher made and tracked the source of feedback that led to revision. The study found that a majority (51.8%) of revisions was actually self/other-induced and teacher and peer-induced changes accounted for only 34.3% and 13.9% respectively. However, when the types of changes were tracked, the majority of self-induced changes were surface level revisions, which had little effect on improving writing quality, whereas peer-and teacher-induced changes were more often meaning-level changes that brought significant changes to writing quality. Overall, the study supported the usefulness of peer and teacher feedback and asserted their positive effect on improving the final draft.

In another study, similarities and differences between teacher and peer feedback were investigated in an EFL setting (Miao, Badger, & Zhen, 2006). After multi-drafting, the improvement of final product over the first drafts was measured, which showed that the

teacher-feedback class achieved significantly higher improvement than the peer-feedback class. Surveys on the usefulness of feedback showed that 98% of students viewed teacher feedback useful or very useful whereas only 40% of students viewed peer feedback useful. In addition, while 90% of teacher feedback was incorporated, only 67% of peer feedback was incorporated in the revision. However, the peer feedback group made more self-corrections than the teacher feedback group, suggesting that peer feedback help students improve their autonomy as writers. In short, while students used peer feedback less than teacher feedback in their revisions, peer-reviewed students tended to be more actively involved in self-correction.

Based on the findings, the authors conclude that both teacher and peer feedback play important roles in EFL students' revision of writing and that the impact of teacher and peer feedback is different. Teacher feedback is incorporated more than peer feedback in revision and leads to greater improvement of subsequent draft but peer feedback brings about a higher percentage of meaning changes and encourages student autonomy. Citing an earlier study by Villamil and De Guerre (1998, p.491), the authors reinstate that "peer revision should be seen as an important complementary source of feedback in the ESL classroom."

Although the studies found peer feedback of L2 learners to be less accurate, less credible and less effective in bringing changes in texts compared to teacher feedback, a study that compared the ratings given by students with that of an instructor indicates that the L2 students' assessment skills are significantly correlated with that of the teachers'. Saito and Fujita (2004) conducted a study that looked into peer writing assessment as an alternative to teacher assessment. The study compared the ratings given by an instructor with the ratings given by 92 Japanese students in EFL writing course and found that the peer and instructor ratings had high and statistically significant correlation (0.72). Contrary to the popular distrust of L2 learners' feedback and ratings, this result suggests that L2 learners are as capable as teachers in evaluating peer writing.

In addition, studies that engaged students in training prior to peer feedback show that the feedback comments of L2 learners improve greatly in quantity, quality and helpfulness upon training. These findings point to the need to re-visit the efficacy of peer feedback of L2

learners suggested by earlier studies that did not involve training. Stanley (1992) provided seven hours of training to college ESL students on peer response procedures and found that the group produced a larger amount of and more specific feedback compared to students who received 1 hour of demonstration. The author concluded that coaching students improves both the quality and quantity of peer feedback. Berg (1999) compared the effect of trained versus untrained peer feedback on revision and writing quality and found that the trained students' feedback led to greater number of meaning changes in the revised drafts, and that trained peer response brought about greater gains between draft 1 and draft 2. In another study, Min (2005) analyzed feedback generated pre- and post- two conference sessions on peer response and found that the comments generated post-training had more relevant and specific comments on global issues. In a following study, Min (2006) investigated the effect of trained peer responses on revision types and quality and found that feedback produced post training was incorporated significantly more during revision and resulted in overall improvement in writing quality. Taken together, ESL and EFL students coached to be successful reviewers were found to generate substantially more relevant and specific feedback comments that were implemented more frequently and resulted in improvement in writing quality.

While the studies on the effect of training on peer feedback point to the potential for high efficacy of L2 learners' feedback comments, it does not offer information on the reliability, validity, and effectiveness of the feedback comments generated by L2 learners. In L1, a series of studies have shown that peer feedback produced by L1 student-writers is highly effective, as reliable and valid as feedback by teachers (Nelson & Schunn, 2008; Patchan et al., 2009). These findings support the use of peer feedback as an alternate source of feedback to student writing. The use of peer feedback as a substitute for teacher feedback will allow students to write more as part of the school curriculum and improve their writing through the benefits of providing and receiving peer feedback. In the L2 community, this opportunity has not yet been explored. It has not been examined whether the peer feedback

by L2 student-writers has similar features and characteristics as the feedback given by L1 counterparts.

To date, no study has systematically compared the feedback comments of L1 and L2 learners produced in a comparable environment. To fill this gap, the present study will compare the feedback comments of L1 and L2 writers, generated under similar conditions. The findings will engender understanding on how these groups perform on producing peer feedback and the efficacy of their feedback comments vis à vis each other and have significant implications on the usage and applications of this popular method of instruction in classrooms with L2 learners.

2.5 LONGITUDINAL EFFECT OF WRITING AND REVISION

Writing process has been explored from a view of general problem-solving activity by many scholars (Hayes and Flower, 1980; Scamadalia and Bereiter, 1986). Insights on teaching and learning of general problem-solving activities can be drawn from cognitive theories about learning (Fredericksen, 1984). According to Anderson's theory of acquisition of cognitive skills, individuals develop a cognitive skill by receiving instruction and practicing the skill. Through repeated practice, a process called knowledge compilation occurs, through which individuals transfer knowledge about the skill into knowledge about how to successfully complete a task involving a skill (Anderson, 1982).

Based on this view, educators have adopted writing programs that emphasize repeated practice of writing skills within a specific task domain such as Writing Across Curriculum (WAC) and Writing in the Disciplines (WID) and the effectiveness of these programs have been found to be positive at the general program assessment level (White, Lutz & Kamusikiri, 1996).

In order to empirically investigate the effect of repeated practice and domain-specific writing on writing development, Johnstone, Ashbaugh and Warfield (2002) examined 279 undergraduate Accounting students who received contextually-relevant writing treatments and 385 business students who had general business writing assignments. Each group of students was divided into three sub-groups based on their progress in the program

(sophomore, junior, senior), which determined the amount of practice they had in writing. When the students were compared on a domain-specific writing task, the mean writing score of accounting students was higher than the other students. Importantly, the writing scores of accounting students increased at progressive points in their curricula ($115.12 < 126.51 < 130.03$), whereas other students' scores remained approximately equal ($114.16 > 109.76 > 108.75$). When the effect of repeated practice was controlled for, writing within a specific domain was associated with superior writing skills.

Regarding the mechanism of how repeated practice of writing helps improve writing ability, Kellogg and Whiteford (2009) explains that repeated practice eases burden on the writer to manage burdensome demands on working memory placed by the task of composition. Composition requires writers to engage in concurrent planning of ideas, generation of text, and reviewing of ideas and text, a task that places heavy workload on executive attention and working memory (Hayes and Flower, 1980; Kellogg, 1996). The heavy demands on working memory can be reduced through practice, supplying the writer with sufficient attention and storage in working memory system that allows the writer to coordinate and control the complex processes of writing.

However, practice in which the writer simply goes through the motions of writing is not sufficient to bring growth. A. Ericsson (2006) proposed that in order for novices to become experts, deliberate practice is necessary. Deliberate practice involves carefully tailored practice tasks and feedback that provides knowledge of results on top of intrinsic motivation and effortful exertion to improve. While Ericsson specified the conditions required for deliberate practice that leads to growth, it is not clear how each component affects writing development and which components are associated with the greatest gains.

With this gap in the body of knowledge, the present study aims to uncover the effect of one of the major components of deliberate practice, or feedback, on writing development in a repeated writing setting. So far, studies on feedback have mainly focused on investigating the immediate effect of feedback on subsequent draft of writing and little has been reported on the long-term effect of feedback on writing development. By investigating

the writing, feedback and revision of students who achieved increase in writing scores in comparison with students who did not during the course of repeated writing and feedback sessions, present study is designed to shed light on the ways in which feedback affects students' writing in the long term during repeated practice.

Extending the theoretical and empirical underpinnings of studies on the development of students' writing competency along with similar findings on other cognitive activities such as chess (Charness, Tuffiash, Krampe, Reingold & Vasyukova, 2005) and musical performance (A. Ericsson, Krampe & Tesch-Romer, 1993), it can be hypothesized that students' ability to provide helpful feedback will improve through deliberate practice. While composition classes that employ the process-approach to writing through multiple drafting using peer feedback typically assigns multiple writing and feedback assignments over a semester, no study has investigated how students' feedback changes over time through repeated practice.

The power law of practice embodies the finding that performance improves as a power function of the amount of practice. This means that the improvement in performance is made rapidly initially and then gradually slows down with more practice. With numerous activities and instructional methods reported as beneficial in improving writing competency of student-writers, understanding how the students' feedback changes over repeated practice and how it affects students' writing development over time will have important implications on optimizing the use of peer feedback in classrooms as well as on writing curriculum development. These inquiries have not yet been explored in the field and will be addressed as part of research questions in the current study.

CHAPTER 3. RESEARCH METHODS

3.1 PARTICIPANTS AND SETTINGS

3.1.1 Setting

The data collection for this study took place in a large university in the U.S. during fall 2010 and spring 2011. Four freshman composition classes, two for native speakers and two for non-native speakers, were recruited to participate in the study. The freshman composition class is a required course. Whereas most incoming freshmen are assigned to ENG 101, non-native speakers of English are eligible to take ENG 101Q in lieu of ENG 101. ENG 101 and ENG 101Q are comparable courses operated and supervised by the same director. The two courses have identical course objectives, curriculum, syllabus, assignments and are taught by instructors who receive the same training. The only differences between the two courses are the target students and class time. ENG 101Q classes are for non-native English speakers with relatively low English proficiency. The classes are offered 50 minutes per session for five days a week, totaling 250 minutes a week. ENG 101 classes are for students with moderately high English proficiency (native or non-native), offered in 75-minute sessions twice a week, or 50-minute sessions three times a week, amounting to 150 minutes a week. ENG 101Q classes cover the same lesson plans as the ENG 101 classes only at a slightly slower pace. Additionally, ENG 101Q classes have 50 minutes of language lessons each week that focused on English grammar, which are absent from the ENG 101 classes.

3.1.2 Participants

3.1.2.1 ENG 101 classes

The two ENG 101 classes recruited were taught by the same instructor in spring 2011. The total number of students registered in the two classes was 36 but only 34 gave consent to participate in the study. The average age of the students was 19.3 years. By gender, the classes had 24 men and 10 women. In terms of first language, 67.5% spoke English, 20.6%

Spanish, 9% Arabic and 3% Chinese. Their SAT scores ranged from 800-1,300 out of 1,600, with around 50% of them falling in the 1,000-1,300 range. Their years of residence in the U.S. or English-speaking countries were 10 years or longer for the majority of students (82%) but some had resided in the U.S. between 4-7 years (9%) and 1-3 years (9%). The data reported in the study pertain to 34 students who gave consent to participate in the study. A summary of students' demographic information is provided in Table 3.1.

3.1.2.2 *ENG 101Q classes*

The two ENG 101Q classes that participated in the study were taught by the same instructor, one in fall 2010 and the other in spring 2011. The total number of students enrolled for the course was 33 but consent was given only by 32. By gender, ten were men and twenty-two were women. Seven native languages were represented in the class: Chinese (13 students or 41% of class), Korean (12 students or 38%), Spanish (3 students or 9.4%), and one student each for Arabic, Finnish, Tamil and Vietnamese. In terms of years of residence in the U.S. or other English-speaking country, 42% of class had stayed less than one year, 35% between one to three years, and 23% four to seven years. The level of their English proficiency was described as intermediate to intermediate-high by their instructor. Their TOEFL iBT scores ranged from 70 to 109 out of 120 with 48% falling in the range of 90-99. Their TOEFL writing score ranged from 21 to 29 out of 30. The data reported in the study pertain to 31 students who gave consent to participate in the study and a summary of students' demographic information is provided in Table 3.1.

Table 3.1 Demographic and Background Information of Participants

	ENG 101 Classes (L1)	ENG 101Q Classes (L2)
No. of students	34	32
Mean Age	19.3	20.5
Gender (M:F)	24:10	10:22
Major	Undeclared (41%)	Science/Engineer (43%)
	Science/Engineer (25%)	Undeclared (18%)

	Social Science (18%)	Social Science (13%), Business (13%)
SAT (Max. 1600)	In range of 800-1300	
TOEFL iBT (Max. 120)		In range of 70-109
First Language	English (68%)	Chinese (41%)
	Spanish (21%)	Korean (38%)
	Arabic (9%)	Spanish (9.4%)
	Chinese (3%)	Finnish, Tamil, Arabic, Vietnamese (3% each)
Years of Residence in English-speaking country	> 10 (82%)	< 1 (42%)
	4-7 (9%)	1-3 (35%)
	1-3 (9%)	4-7 (23%)

3.1.2.3 Instructors

Instructors for both classes were native English-speakers who were graduate students in doctoral programs in the English department of the university. The instructor for ENG 101Q had taught the same course for two semesters prior to fall 2010 and had experience teaching Spanish and Portuguese as a second language. The instructor for ENG 101 had taught classes that were the same or similar to ENG 101 for three years before participating in the study.

3.2 RECRUITING AND TRAINING

3.2.1 Instructors

The instructors were recruited through personal contact. They were given a 30-minute training session by the researcher on how to set up and use the online peer review tool SWORD.

3.2.2 Students

At the beginning of the semester, all students in the classes of the two recruited instructors were informed of the details of the study by the researcher during the class and asked to express their intention to participate (or not) in the study by writing (not writing)

their names on a background information survey. The procedures for the study, that is, conducting peer review in groups of four using the online tool SWORD, were adopted as the normal procedures for the class, and therefore all students were required to engage in the peer review activity in the same manner regardless of their participation status in the study. What participation in the study dictated was which students' data were included for analysis. In ENG 101 classes, 34 out of 36 students agreed to participate and in ENG 101Q classes 32 out of 33 students gave their consent to participate in the study.

Approximately a week before the first peer review session, a lesson on peer review was held in class by the instructors in all four classes using PowerPoint slides. The session addressed the purposes and benefits of peer review and how to provide peer feedback that is useful to the writer based on findings from recent studies on peer review. Some of the guidelines shared with the students included; "provide feedback that is detailed and specific," "try to provide solution to problems rather than just pointing out problems," "try to provide explanations on why something is a problem and why certain solutions are recommended." Examples of good and bad review comments were included (see Appendix D). The examples were compiled by the researcher from students' actual peer feedback comments from an ENG 101 course. Other examples came from a tutorial on peer review posted on a webpage *peerfeedback.net* (<http://www.pitt.edu/~schunn/feedback/>). Students were also told about recent peer review findings suggesting that there is more to benefit from giving feedback than receiving feedback (Lundstrom & Baker, 2009) to motivate them to maintain rigor in providing multiple reviews throughout the semester.

3.3 INSTRUMENTS

3.3.1 Writing Assignments

As part of the course requirements, all students in all four classes were required to submit three academic essays over the semester. Students were required to write three drafts for each essay. The first draft was not reviewed by the instructors and was used solely for

providing and receiving peer feedback. The second draft was a revision of the first draft addressing the peer review comments that the students received. The second draft was submitted to, reviewed, and commented on by the instructor. The third draft, which was the final draft, was a revision of the second draft incorporating the comments of the instructor. Grades were given on the second and the final drafts of paper but not on the first drafts.

The class procedures of the present study were different from other classes in SWORD studies in two aspects. In previous studies that used SWORD, part of the final course grade the students received was based on the helpfulness of their feedback rated by the students who received their feedback. In addition, in earlier SWORD studies, students were asked to rate the essays on three dimensions of feedback prompts on a scale of 1 to 7 in addition to providing written feedback comments. These ratings were compared against the mean of the group that reviewed the same essay. The ratings' degree of divergence from the group means was used as a measure to reflect accuracy of the rating and students' grades were penalized for inaccuracy. These were measures put in place to motivate students to take peer review seriously and to work towards providing more helpful feedback. However, in the current study, the instructors did not feel comfortable using these measures as part of students' course grades, and thus while students were asked to rate the essays they reviewed and back-evaluate the helpfulness of feedback they received, these were not counted toward their final course grades. As such, students in the present study may have been less motivated to take peer review and feedback seriously compared to the student- participants in other SWORD studies, and this may have influenced students' behavior and performance on peer review and feedback.

The characteristics/genres of essays required over the semester varied slightly from the first assignment to second and third. The first assignment was more of a summary, the second an analysis, and the third more of an argumentative essay. Since each genre of a writing requires different genre knowledge and skills to write, this difference in the genres of three assignments may have affected the performance/grade students achieved on each

assignment. The instructions given by the instructors for the three writing assignments are listed below.

Class ENG 101

Assignment 1: Write an essay outlining the positions of three different articles.

Assignment 2: Rhetorically analyze a text that presents an argument.

Assignment 3: Make an argument for a particular position in a controversy.

Class ENG 101 Q

Assignment 1: Write an essay of 5-7 pages outlining the positions of three different articles.

Assignment 2: Write a 5-7 page essay rhetorically analyzing two different sides of your controversy. Compare and contrast the way both sides argue.

Assignment 3: Write a short paper attempting to convince a specific audience of your point of view on your chosen controversy. Papers should be at least 2 pages.

Although the instructions given for the assignments were the same for the ENG 101 and ENG 101Q classes, the length requirement for the first draft of essays was different. Whereas ENG 101Q classes were asked for a full length paper (5-7 pages) for the first draft, the same length required of the final draft of the essay, ENG 101 classes were asked for a minimum of three pages, which is only half the length of the final draft. The difference in the length requirement of essays could have affected the degree of completeness the students aimed for when preparing their first draft of paper.

3.3.2 SWORD

As a tool to aid students' peer review, a web-based peer review system SWORD; Scaffolded Writing and Re-writing across Discipline (Cho & Schunn, 2007) was used throughout the semester. SWORD has been used in many different courses and disciplines ranging from English and education to medicine and law, in more than a dozen universities in the U.S and abroad. SWORD is reported to be effective in providing the conditions for peer review that enhance the quality of students' review comments. In addition, the feedback comments produced by students through SWORD are found to be as reliable and valid as instructor feedback (Patchan et al., 2009).

SWoRD was used in this study for two purposes: first it served as a means to secure a similar setting and environment for the peer review sessions for all students across the four participating classes. By using SWoRD, the students were geared towards doing peer review in the same format and manner using the same prompts, which is expected to reduce variability in the peer review sessions across classes and thereby increase comparability of the data generated from the peer review sessions.

Secondly, SWoRD was developed specifically to aid students' peer review activities based on the findings on best practices for writing instruction (Hillocks, 1984) and offers unique features such as automatic random distribution of multiple papers, back-evaluation (ratings of the helpfulness of the feedback from the author back to the peer-reviewers), and a focus on global/high-level features of the reviewed paper. Thus, using the tool repeatedly over a semester was expected to have some training effect on the students in developing skills to provide more helpful feedback.

SWoRD has four steps in one cycle: 1) composition, 2) peer review and rating, 3) back-evaluation, and 4) revision. Students write a paper and submit it through SWoRD. SWoRD automatically assigns the paper to the number of reviewers pre-set by the instructor. For this project, each student was asked to review four peers' papers because a study reported that four reviews are the minimum number required to obtain an acceptable level of reliability and validity in peer review (Cho et al., 2006).

SWoRD allows instructors to provide detailed and structured rubrics/questions that lead the students through their peer review process. The instructors of two classes worked together to design mutually agreeable rubrics/questions to be used for peer review in both classes and came up with the following, used throughout the semester for all three writing assignments in all four classes:

1. Global Organization - Does the paper have a clear thesis? Is each paragraph's relationship to the thesis clear? Does the paper have a well-structured introduction and conclusion?

2. Paragraph Organization - Is each paragraph organized around one central idea? Are these ideas stated in topic sentences? Do the paragraphs have clear and effective transitions? Do the paragraphs have enough detail?
3. Sentence-Level Comments - Are the sentences clear and well-formed? Do the sentences follow from one another? Is there good use of sentence variation? Are grammar and punctuation correct?

For each paper reviewed, students were asked to write comments on the areas/points specified by the rubrics/questions above. The rubrics used in the current study were prepared by the instructor-participants of the study. These rubrics differ from those used in previous SWoRD studies, which asked about the originality of the ideas (content) and the use of well-supported claims. Instead, the rubrics in the present study focus more on structure, the well-formedness of the paper at global, paragraph, and sentence levels. This difference may have discouraged students in the study from commenting on the content. When students finished writing comments for all rubrics/questions, they were asked to give a rating on a scale of one to seven on each of the rubric/question on how good the paper was in terms of those questions.

After receiving peer feedback, the students were asked to revise and submit a second draft of their paper by incorporating the peer feedback they had received. After completing their revision, the students were asked to rate and comment on the helpfulness of the peer feedback they had received. By evaluating the feedback they had received, the students not only provided useful first-hand information on how and why the feedback givers' comments were helpful or not helpful but by doing so, they could reflect upon their own feedback that they had provided to other students and learn how they could improve their feedback.

3.3.3 Student Background Survey

At the beginning of the semester, a background survey was administered to all students in all four classes to collect information about some of the individual variables that

may affect the performance of their peer review. The survey was constructed based on a survey used on previous peer review studies (Charney et al., 1995; Patchan et al., 2009). The survey consisted of 32 items, with five multiple choice questions and seven open-ended questions on basic demographic information, standardized test scores (TOEFL and SAT), past peer review experience, and language background. In addition, 20 likert-scale questions on perceptions and beliefs about learning and writing were included in the survey. A full version of the Student Background Survey is provided in Appendix A.

3.3.4 End-of-Semester Student Survey

At the end of the semester, students completed a survey about the experience they had had with peer review and writing throughout the semester. The survey was based on the survey in a previous peer review study (Kaufman & Schunn, 2010). Around 20 items were borrowed and/or modified from Kaufman et al's study and around 40 new items were added. The survey is comprised of three sections; 1) reflection on peer review experience, with three open-ended, five multiple-choice, and 24 likert-scale questions, 2) reflection on writing experience, with twenty likert-scale questions, and 3) reflection on using SWoRD, with ten likert-scale questions. A full version of the End-of-Semester Student Survey is provided in Appendix B.

3.3.5 Instructor Survey

At the end of the semester, both instructors completed a survey that included questions on 1) demographic information, 2) past teaching experience, 3) beliefs and attitudes towards peer review, 4) goals and priorities for the course, 5) instructions given and amount of time spent in class related to peer review 6) observations and reflections on students' improvement in writing over the semester, and 7) experience regarding using SWoRD. The survey consisted of 13 open-ended questions and 19 likert-scale questions. A full version of the Instructor Survey is provided in Appendix C.

3.4 PILOT STUDY

3.4.1 Need for Pilot Study in ESL class

Although SWORD has been reported effective for peer review in multiple classes and settings at various institutions, its use in L2 classes has not yet been reported. Because SWORD was designed to elicit peer review feedback in a form of full sentences on a feedback form as opposed to inserting editing symbols or fragment phrases on the margin of essays, it was possible that the performance of peer review mediated by SWORD could be affected by the level of writing fluency of the reviewers in L2 setting. Thus, before using SWORD in a non-native freshman composition class in the main study, it was necessary to test SWORD's viability as a tool that facilitates peer review for the non-native English users. To this end, a pilot study was conducted with a purpose of testing whether the students at an intermediate proficiency level of English using SWORD generate useful peer feedback comments in narrative sentences that are comprehensible to other students in a reasonable time period.

3.4.2 Setting and Participants

The pilot test was conducted in summer 2010 in an ESL class of a language center affiliated with a large southwest university in the U.S. The class was an intermediate writing class comprised of 14 non-native English speakers. A few of the students had admission to the university (undergraduate and graduate) and were enrolled in the class to improve their writing skills before their university program started in fall. The majority of the students were taking the course to advance their English language skills required for admission to universities in the U.S. So it was reasonable to assume that the English proficiency of the students in this course was at par or lower than that of the students who enrolled in the ENG 101Q class in the fall.

The instructor of this course was a native English speaker who had over 12 years of experience in teaching writing in ESL classes. The instructor had no prior experience in using SWORD and did not feel confident in administering the tool in class so the peer review session using SWORD was led by the researcher.

Of the 14 students enrolled in the class, 13 were men and one woman. The first language of a majority of the students was Arabic with a few speaking other languages including Thai and Korean. Most of them had never reviewed more than one peer paper during peer review. None of them had experience using SWORD prior to the pilot study.

3.4.3 Procedures

A brief training session was provided to the students a week prior to the peer review session. Purposes and benefits of peer review were shared with the students and how to use SWORD was demonstrated. Students were assisted in making SWORD accounts and uploading their drafts to SWORD.

The number of reviews was set at three considering the work load of students and time allowed for peer review as all reviews had to be completed within the 90-minute class time. The peer review session was held in a computer lab for 90 minutes.

3.4.4 Results

During the 90-minute session, most students managed to review and provide peer review comments for three essays that were one to two double-spaced pages in length. During the review of the first paper, many students displayed unfamiliarity with answering rubric questions in a separate window rather than marking their comments on the essay as they read it. However, from the second paper onward, students demonstrated comfort in using SWORD.

The pilot study showed that students at an intermediate level of English who have not yet acquired admission to U.S. universities were able to provide a fair quantity of helpful comments in comprehensible sentences in a reasonable time. In a survey conducted at the end of the peer review session, a majority of the students reported that the feedback they received helped them identify problems in their writing and improve their writing through revision.

The positive results gained from the pilot study removed concerns around adopting SWORD as a peer review tool in a non-native English class and supported the selection of SWORD for the main study.

3.5 DATA COLLECTION

3.5.1 Data from SWORD: Essays, Feedback Comments, and Back-evaluation

Students' essays and peer feedback comments were collected from the three peer review sessions using SWORD. SWORD has a database that stores all the essays that students submit and all the feedback that a particular essay receives.

3.5.1.1 Peer Review Session One

The first peer review session was held approximately five weeks after the semester had started. Students were introduced to SWORD and were assisted in registering accounts in SWORD prior to the peer review session. The ENG 101 instructor allowed students to work on peer review during the regular class time outside their regular classroom because their classroom was not equipped with computers. ENG 101Q classes were equipped with computers, which allowed the instructor to do peer review during class time. In the interest of time, the instructor asked the students to conduct one or two reviews in class and the rest as homework.

For a variety of technical reasons, a number of students were not able to participate in peer review during the first session. Of the total 66 participating students, only 57 (29 in L1, 28 in L2) students submitted their essays on time. But even for the students who successfully uploaded their writing on time, the completion rate of providing feedback on all four assigned essays stood at 80.6% for L1 classes and 76.2% for L2 classes.

3.5.1.2 Peer Review Session Two

Approximately four weeks after the first peer review session, the second peer review session was held. Extra effort was made to encourage students to upload their essay before the deadline and solutions for tackling technical problems were communicated to the students.

Despite the increased precautions, the number of students who turned in their drafts on time remained the same as the first session. Out of 66 participating students, only 57

(L1=28, L2=29) submitted their essays before the deadline. In terms of completion rate of providing feedback to four assigned essays, the rate decreased from 80.6% to 67.5% for L1 students while the rate increased from 76.2% to 93.5% for L2 students.

3.5.1.3. Peer Review Session Three

Approximately four weeks after the second peer review session, the third peer review session was held. Having already performed peer review using SWORD twice, few technical issues were expected. However, the number of students who uploaded essays before the deadline dropped to 53 (L1 = 27, L2 = 26). The feedback completion rates were 80.6% for L1 students and 84.3% for L2 students.

3.5.1.4 Back Evaluation

After receiving peer review comments, the students were asked to rate the helpfulness of each reviewer's feedback on a scale of one to five, with one being the least helpful and five being the most helpful. Although students were encouraged to provide back evaluation, it was not a mandatory activity that was tied to the course grade system. Although 54.4% of all comments received back evaluation and rating in the first peer review session, the figure went down to 46.9% in the second session and further down to 24.6% in the third session.

3.5.2. Non-SWORD Data: Instructors' Grades and Survey Questionnaires

3.5.2.1 Instructors' Feedback and Grades

Throughout the semester, both instructors provided the researcher with the grades and the written comments they provided to each student's second and third drafts of the writing assignments. Aside from the written feedback, the instructors held two face-to-face conference sessions with the students during the semester to provide oral feedback, following the second draft of the first assignment and the second draft of the second assignment. The oral comments were not recorded.

3.5.2.2 Questionnaires

The Students' Background Survey was administered in class at the beginning of the semester and was completed by 66 participants. The Students' End-Of-Semester Survey was

administered in class for three classes and via email for one class. Only 55 students filled out the survey, with the lowest response rate coming from the email channel. The Instructor Survey was administered via email after the end of classes.

3.6 DATA ANALYSIS

3.6.1 Part 1: Analysis of Feedback Comments

Students' feedback comments were analyzed for number and length in order to address three questions. First, the general participation level of L1 and L2 students was compared by looking at the number of essays students reviewed during the semester, the total number of idea units generated per review, and the number of words produced per review and per idea unit. In addition to the language group effect, a gender effect on these variables was also investigated.

Second, L1 and L2 students' comments were compared by type: by their function (praise, criticism, reflection), scope (macro/meaning-level, micro/surface-level), and specificity (general and specific). Criticism comments were further divided by content (problem-only, solution-only, or problem+solution).

Third, longitudinal changes were investigated by comparing the feedback profile of each language group across the three peer review sessions.

3.6.1.1 Coding Scheme for Feedback Comments

The coding scheme was a modification of a scheme used in a previous peer review study (Patchan et al., 2009). In Patchan's study, function was an important dimension that differentiated the comments of subject-matter experts and writing-instruction experts. In addition to Patchan's study, other studies have also found scope to be important (Ferris, D., 1997; Hansen, J., & Liu, J., 2005; Miao, Y., Badger, R., & Zhen, Y., 2006). Specificity was found both by Patchan et al. (2009) and Cho et al. (2006) to be crucial in the writer's decision to implement feedback during revision. The definition of feedback categories and the examples from students' feedback are provided in Table 3.2.

Aside from the sixteen categories reported in the present study, the researcher had three additional categories that were initially included in the coding scheme but later removed during the quantitative analyses process due to low instances of occurrences of feedback in these categories. These categories include mitigation language, summary, and incomprehensive comments. These three categories combined account for the gap between the whole (hundred percent) and the sum of praise, criticism, and reflection categories reported in the results section.

Table 3.2 Definition and Examples of Coding Categories

Coding Category	Definition	Example
1. Praise	When a feature in a writing is stated positively	
A. Praise-general	Without reference to a) a specific location, b) reason why or, c) a detailed account of the problem (how)	<i>Good use of sources and good organization.</i>
B. Praise-specific	With reference to a) a specific location, b) reason why or, c) a detailed account of the problem (how)	<i>Each body paragraph has a good main idea that is backed by evidence from their sources.</i>
2. Criticism	When a feature in a writing is stated negatively	
A. Meaning-level	Pertaining to one or more sentences that would bring change in meaning	
i. Problem-general	Describes problem without reference to a) the specific location, b) reason why or, c) a detailed account of the problem (how)	<i>There is really a big transition from paragraph to paragraph.</i>
ii. Solution-general	Suggests changes without reference to a) the specific location, b) reason or, c) a detailed account of the solution	<i>I think just avoiding repeating yourself should help.</i>
iii. Problem-specific	Describes problem with reference to a) the specific location, b) reason why or, c) a	<i>Each paragraph seems summary of the history of DDT</i>

	detailed account of the problem (how)	<i>rather than focusing on it as a controversial issue.</i>
iv. Solution-specific	Suggests changes with reference to a) the specific location, b) reason or, c) a detailed account of the solution	<i>Maybe cite some of the sources that you used to find this information so we know that you are not just making it up.</i>
v. Prb+Sol-general	Describes problem and suggests changes without reference to a) the specific location, b) reason, or c) a detailed account of either the problem or the solution	<i>Your paragraphs are a little bit jumbled with information that should be somewhere else. Make a separate paragraph.</i>
vi. Prb+Sol-specific	Describes problem and suggests changes with reference to a) the specific location, b) reason, or c) a detailed account of either the problem or the solution	<i>The 2nd & 3rd paragraph kind of repeats itself, so maybe a different idea to elaborate on would be better to suit the opening paragraph.</i>
B. Surface-level	Involves a change within a sentence that does not bring change in meaning	
i. Problem-general	Describes problems without reference to a) the specific location, b) reason why or, c) a detailed account of the problem (how)	<i>There are some grammar and punctuation that needs to be corrected.</i>
ii. Solution-general	Suggests changes without reference to a) the specific location, b) reason or, c) a detailed account of the solution	<i>In the first paragraph some sentences could be combined to create longer ones</i>
iii. Problem-specific	Describes problems with reference to a) the specific location, b) reason why or, c) a detailed account of the problem (how)	<i>There are a lot of times where you use improper tenses, suffixes, etc. For example, you use leaves the reader wonder as opposed to ... wondering</i>

iv. Solution-specific	Suggests changes with reference to a) the specific location, b) reason or, c) a detailed account of the solution	<i>I think you should use a different word than aspire, I'm not sure that is the right word to use. I aspire to reconfirm to the audience that a tax on carbon emissions...maybe use a comma after aspire.</i>
v. Prb+Sol-general	Describes problem and suggests changes without reference to a) the specific location, b) reason, or c) a detailed account of either the problem or the solution	<i>There were a couple times where you used the wrong words (... States of American as opposed to ... States of America)</i>
vi. Prb+Sol-specific	Describes problem and suggests changes with reference to a) the specific location, b) reason, or c) a detailed account of either the problem or the solution	<i>I only see a few transition words. Try to use more transition words such as then, but, although, therefore, etc. These words would make your sentences run more smoother.</i>
3. Reflection	Describes how the reviewer benefited in his/her writing from reviewing the particular paper	
i. General	Without reference to a) the specific location, b) reason, or c) a detailed account of the benefit	<i>Look more into my organizing of ideas and topics.</i>
ii. Specific	With reference to a) the specific location, b) reason, or c) a detailed account of the benefit	<i>I completely forgot to cite my quotes, which means that I am going to work on it.</i>

Each feedback comment was divided into an idea unit, defined as a chunk of feedback that pertains to a single feature of writing regardless of a sentence boundary. In this sense, a sentence can be divided into several idea units, or several sentences can be combined into a

single idea unit. The feedback comments from the sixty-five participants over three sessions of peer review amounted to 4,227 idea units. Once the feedback comments were separated by idea unit, then each idea unit was categorized into one of sixteen categories defined in Table 3.2. The researcher divided the feedback comments into idea units and coded all 4,227 idea units generated by 65 participants in three peer review sessions.

To check coder reliability, a second coder was recruited and coded approximately 10% of the total comments, or 420 idea units. The second coder was a sixth year doctoral student in the department of English of the University and had taught ENG 101 for four semesters prior to coding the comments. After the two coders shared the definition of categories and examples, the two coders achieved over 90% agreement in total number of idea units and in all of the categorical breakdowns.

3.6.1.2 Participation Level by Language Group and Gender

In the first pass of analysis, the goal was to gain a broad sketch of L1 and L2 groups' overall feedback production and not session-specific characteristics. The students were first grouped by language (L1 or L2) then by gender. All data available from 59 participants were used in the analyses, whether a student participated only in a single peer review session or all three sessions over a semester. The variables on which the groups were compared are as below.

- Total number of essays reviewed per student during the semester
- Number of idea units generated per student per review
- Number of words generated per student per review
- Number of words generated per student per idea unit.

3.6.1.3 Comparison of Feedback Profiles of L1 and L2 Groups

3.6.1.3.1 Missing Data

The students participating in the study submitted three writing assignments and conducted three peer review sessions during the semester, and not all participants generated

feedback data at all three sessions. While the total number of participants was 66, seven of them did not participate in any feedback sessions and therefore no feedback comments were available from those students. Of the remaining 59, the number of students who participated in all three sessions of peer review was only 32, and the rest participated in only one or two sessions.

Instead of replacing mean value for the missing data, all missing data were treated as list-wise deletion in order to secure accuracy in the analyses.

3.6.1.3.2 Statistical Technique: MANOVA

Deciding the appropriate technique for statistical analysis is best made on the basis of the research question (Tabachnick & Fidell, 2001). Research questions one and two pertain to the peer feedback profiles of the two language groups (L1 and L2) and how each group's feedback changes over time. These research questions fall into *the significance of group differences* category under Tabachnick and Fidell's three categories of research questions. For research questions that fall under this category, the choice of technique then "hinges on the number of independent variables and dependent variables" (p.21).

Multivariate analysis of variance, or MANOVA, was chosen as the statistical technique for these research questions as the method evaluates differences among centroids for a set of dependent variables when there are two or more groups (Tabachnick & Fidell, 2001), and is appropriate when the main interest of the study is to investigate the effect of grouping variables on a collection of multiple outcome variables (Keselman et al., 98). A statistical software package SPSS version 19 was used to run MANOVA. Prior to executing MANOVA, the data's normality was tested by checking the frequency chart, histogram and skewness of data for each dependent variable.

3.6.1.3.3 Middle-level Coding Scheme

Instead of looking at the distribution of feedback in the sixteen low-level categories, the categories were collapsed across boundaries to create more meaningful middle-level categorical units. These categories each include a set of variables that represent 1) feedback

types, 2) functions of feedback, 3) degree of specificity and 4) scope of the writing feature the feedback is addressing. MANOVA was run separately for each of the four middle-level categories. The detailed sub-variants that form each middle-level category are listed in Table 3.3.

Table 3.3 Middle-level Categories and Their Sub-variables

Middle-level Variables	Sub-variables
Types	Praise Criticism Reflection
Function	Problem Solution Problem+Solution
Specificity	General Specific
Scope	Meaning-Level Surface-Level

To compare feedback profiles of the two groups, data from all three peer review sessions were combined. As for the analyses on the longitudinal effects, a within-subject repeated measures design was used.

3.6.2 Analysis of Effect of Peer Feedback and Revision on Writing over Time

3.6.2.1 Analytic Framework

In the literature of writing studies, few have looked into the topic of the long-term effect of peer review on writing development giving little guidance for research question three: whether students change in the ways they provide peer review and revise papers over time. Therefore, question three was investigated based on grounded theory (Glaser & Strauss, 1967) and inductive strategy. Unlike deductive strategies, inductive strategies include a thorough examination and analysis of all data. In other words, researchers try to find theories to explain all the data they obtain from the studies (Merriam, 1998).

A case study is a research design widely adopted in qualitative research that provides intensive descriptions and analyses of a single unit or bonded system (Smith, 1978), such as

an individual, program, group, intervention, or community (Merriam, 1998). Due to the vast amount of data to be investigated as well as the exploratory nature of the present study, a case study design was employed to uncover the factors/variables involved in how students' peer review activities affected long-term growth in their writing ability.

3.6.2.2 Sample Case Selection

The sample case selection followed *purposeful strategy* (Patton, 1990), which is the most common form of nonprobabilistic sampling strategy. Chein (1981) notes that in order for the investigator to discover, understand, and gain insight, (s)he must select a sample from which most can be learned, and thus nonprobabilistic sampling is the most appropriate method of choice for most qualitative research.

Sample students for the case study were first screened for completeness of data set. Students who participated in fewer than two peer review sessions were eliminated. Next, students who did not complete the background survey and the post-semester survey were excluded. To heighten the contrast between classes, students in L1 classes whose first languages were not English were excluded even though they have resided in the U.S. for most of their lives.

According to the student survey, the most prominent and common goal pursued by participating students was achieving good grades, of which a large portion was determined by the grades students received on their writing assignments. As the course objective and lessons were centered around understanding good writing and helping students write better throughout the semester, it seemed reasonable to take the grades students received on their writing assignments as a measure that reflect students' writing competency. Of the three drafts that students wrote for each assignment, the grade that instructors gave on the second draft seemed to best represent students' general writing ability because the first draft was often an incomplete version of writing with less detail and shorter length compared to the second and the final drafts. The grades on the third drafts were not considered as a good representation of students' writing ability since the grades on the third drafts were given

primarily based on how much of the instructors' feedback on the second drafts were incorporated during revision rather than based on general aspects of writing quality.

Students whose grades on the second draft consecutively increased across the three assignments were selected as an 'improve group'. On the other hand, the students whose grades remained approximately the same across assignments, with one or less notch increase from the first to the second or the third were selected as a 'non-improve' group.

This selection process resulted in four students in the improve group and two students in the non-improve group in L1 classes and four students in high and three in low groups in L2 classes. However, it needs to be noted that the majority of students in L2 group received scores clustered around 85-95 and the difference in scores between the 'improve' and the 'non-improve' group was slim. As such, the sample students selected for each group may not constitute good contrast groups in terms of degree of improvement in their writing scores over a semester. The background and writing scores of the sample cases are provided in Table 3.4.

Table 3.4 Background and Writing Scores of Sample Cases

[L1 Group]

No	Grp	Name	L1	yrs in US	sex	age	major	SAT (writing)	writing score
1	High	Ethan	Eng	> 10	M	19	Science/Engr	640 (12)	F/B-/A-
2	High	Doug	Eng	> 10	M	*	Science/Engr	520 (8)	B+/B+/A
3	High	Peter	Eng	> 10	M	20	Humanities	600 (10)	D+/C-/B+
4	High	Jordon	Eng	> 10	M	18	Business	550 (10)	C-/C/B
5	Low	Wendy	Eng	> 10	F	18	Undeclared	680 (10)	B-/C+/B-
6	Low	Amy	Eng	> 10	F	18	Social Science	570 (8)	B+/B+/B+

[L2 Group]

No	Grp	Name	L1	yrs in US	sex	age	Major	TOEFL/ SAT (writing)	writing score
1	High	Megan	Chinese	<1	F	20	Science/Egnr	98 (25)	92/95/94
2	High	Harry	Finnish	<1	M	19	Undeclared	110 (29)	85/94/94
3	High	Steve	Tamil	<1	M	19	Undeclared	98 (28)	92/94/95

4	High	Sally	Korean	<1	F	23	Humanities	89 (21)	90/91/93
5	Low	Yves	Chinese	1-3	F	20	Undeclared	90 (*)	86/78/75
6	Low	Jennie	Korean	4-7	F	19	Science/Egnr	97 (22)	94/93/90
7	Low	Heather	Chinese	4-7	F	21	Science/Egnr	580 (9)	94/89/90

* : *not reported*

3.6.2.3 Themes and Sub-themes

For the analyses of the case study, the constant comparative method of data analysis (Glaser and Strauss, 1967) was used. In the constant comparative method, the researcher begins with a particular unit of data and compares it with another unit of data in the same or different data set until these comparisons lead to tentative categories (Merriam, 1998). Because the basic strategy of the constant comparative method is compatible with inductive, concept-building orientation of qualitative research, the method has been adopted widely by many qualitative studies. The list of data analyzed in constant comparative method in a recurring manner during the preliminary phase is as below.

- Students' demographic information
- Students' past peer review experience
- Students' belief and attitude towards peer review and writing
- Feedback comments received by students
- Feedback comments produced by students
- Students' first and second drafts of essay for all three assignments
- Instructors' written feedback to student's essay
- Grades awarded to each draft of essay by the instructors
- Ratings on first draft given by peers
- Back evaluation rating and comment on helpfulness of feedback
- Types/amount of revision made between the first, and the second drafts
- Incorporation of feedback during revision

After recursive process of reviewing, reflecting, connecting, and meaning-making of the data, five themes emerged as factors that separate the improve and the non-improve

groups, suggesting that how students engage in these five themes may influence the direction and the degree of their writing development over a semester. Some of the themes were broken down into a set of sub-themes for more detailed analysis. These themes and sub-themes were used as the points of analyses for comparing the improve and the non-improve groups. The data of improve and non-improve groups on these themes and sub-themes were counted/coded and compared. The list of five themes and their sub-themes is provided in Table 3.5.

Table 3.5 Themes and Sub-themes

Theme	Sub-themes
1.Types of Revisions	<ul style="list-style-type: none"> a) Changes in surface features such as punctuation and grammar b) Changes in style/expression c) Extension/reduction of existing idea (in part) d) Addition/deletion of new idea (in whole) e) Change in organization: relocation/shifting of texts across sentence or paragraph boundaries
2. Number of Revisions	Changes that constitute a single type of revision were counted as one revision regardless of the amount and/or the magnitude of changes made to the text.
3. Feedback Received	<ul style="list-style-type: none"> a) Number of criticism comments received (in idea unit) b) Number of global-level criticism feedback received
4. Feedback Provided	<ul style="list-style-type: none"> a) Total number of feedback comments generated (in idea unit) b) Proportions of praise, criticism, and reflection comments c) Percentage of macro vs. micro comments d) Percentage of general vs. specific comments e) Percentage of problem vs. solution vs. problem+solution comments
5. Incorporation	<ul style="list-style-type: none"> a) Percentage of criticism feedback incorporated b) Percentage of global criticism feedback incorporated

3.6.2.4 Theme 1: Types of Revisions Made

Revision is a sequence of changes to any existing text at any point during the writing process (Flower et al, 1986; Hayes & Flower, 1986; Sommers, 1980). Most of earlier studies that investigated revisions, both in L1 and L2 literature, adopted the taxonomy of revision strategies by Faigley and Witte (1981) that categorizes revision into ones that affect meaning versus ones that do not (Ferris, D., 1997; Hansen, J., & Liu, J., 2005; Miao, Y., Badger, R., & Zhen, Y., 2006). However, in order to capture the breadth and depth involved in revision, a more sophisticated coding scheme seemed necessary. For example, the scope of revision showed much diversity ranging from fixing a spelling or replacing a single word to adding or removing several sentences to re-organizing or adding several paragraphs.

Some of the initial data analysis revealed that some students employed a wider range of revision strategies while others used only limited number of revision strategies. The data seemed to indicate that different types of revision require different levels of skills and efforts on the part of the writer for implementation. Furthermore, the effect each revision type has on the subsequent draft is substantially distinct and is crucial in determining the quality of writing resulting from the revision.

In order to delve into the different types of revision strategies employed by student writers and to investigate their effect on the writing quality as well as the development of students' writing skills, the researcher devised a typology of revision types expanding on the widely adopted Faigley and Witte's (1981) category of meaning vs. surface-level revision types. Meaning and surface categories were divided into two sub-categories in order to better delineate the wide range of changes that can be made to texts from different types of revision strategies. The sub-categories were developed based on the revision strategies found in the empirical data of students' actual revision works and the effect such changes bring to the subsequent writing.

In addition, a new category of *organizational change* was added. Students' revision data showed that organizational changes, primarily cutting & pasting a paragraph(s) or parts of a paragraph(s) was one of commonly employed revision strategies that entailed changes

that affected the writing as a whole beyond changes in meaning, style, or form of local part(s) of a paper. Apparently, the prominent use of this type of revision strategy can be attributed to the wide use of computer word programs as the mode of composition. This may explain why the category was not included in Faigley and Witte's (1981) original taxonomy of revision strategies when it was first developed in the 1980s. Table 3.6 provides the definition and examples of the revision strategies used in this study.

Table 3.6 Typology of Revision Strategies

Revision Strategies	Definition	Examples from Appendix E
<u>Surface-Level</u>		
Type 1: Punctuation/Grammar	Change in forms such as punctuation or grammar that brings no change in meaning.	R # 5, 6
Type 2: Style/Paraphrasing	Change in word(s), phrase(s), clause(s) or sentence(s) that may affect the impression on reading without change in meaning.	R # 2, 4, 11, 14, 18
<u>Meaning-Level</u>		
Type 3: Extension (Reduction) of Existing Idea	Elaboration/extension (or reduction of) of an idea already presented, typically by adding (removing) a sentence(s) in part or in full.	R # 7, 8, 10, 12, 13,15
Type 4: Addition(Deletion) of New (Whole) Idea	Addition/deletion of a whole chunk of idea(s) in writing, typically by adding/deleting a paragraph(s) in part or in full.	R # 3, 9
<u>Type 5: Organization Change</u>	Shift of a paragraph(s) or parts of paragraph from one place to another in writing.	R # 19

An example of how the number and types of revisions were counted and categorized according to the types and definitions in Table 3.6 is provided in Appendix E. Each occurrence of a change in the text from the previous version to the current version was counted as one revision regardless of its scope as long as the change constituted for a single revision strategy. For instance, a deletion of a single word and the re-location of a whole paragraph each counted as one revision and were categorized under strategy Type 2 (style/paraphrasing) and Type 5 (change in organization) respectively. When there were

multiple occurrences of the same revision across the paper, as was often the case with format and grammar corrections, all the same changes were collectively counted as one revision instead of counting each occurrence as one revision. For example, a change of placing a page number after a citation or changing a past tense verb to present tense might take place multiple times in a single paper. In such cases, all changes of adding a page number after a citation were collectively counted as one revision, and all changes of a past tense to a present tense were counted as another revision.

3.6.2.5 Theme 2: Number of Revisions Made

The number of revisions made was counted by first identifying the changes from draft 1 to draft 2 using document comparison function in Microsoft Word. For each change identified, the type of revision was tagged where each revision type was counted as a single revision. As such, the degree of change to the text resulting from a single revision varied widely. For instance, Type 1 revision usually entailed changes in a word or a number of words in but Type 4 revision typically entailed changes at paragraph level. The number of revisions made per student per draft was counted for each revision type and then summed.

3.6.2.6 Theme 3: Feedback Received

In regards to the feedback received by students, the following sub-themes were coded and counted for comparison between the improve and the non-improve groups.

- Total number of feedback comments received (in idea units)
- The number of criticism comments received (in idea units)
- The number of global vs. paragraph vs. sentence-level feedback received
- The feedback comments received by a draft of paper were categorized into global vs. paragraph vs. sentence-level feedback.

Each feedback comment was broken down into idea units and the number of criticism comments was counted. An example showing how the feedback comments were categorized and counted is illustrated in Table 3.7.

Table 3.7 Example of Feedback Received, its Categorization and Counting

Comments for Steve's draft1S1

Reviewer	Global Organization	Paragraph Organization	Sentence-level Comments
Penny	Actually, your thesis is really good. Your introduction is well-organized and your suggest author's main ideas very well. However, <i>conclusion is unclear [1]</i> .	<i>Your topic sentences of each paragraph are not clear [1]. And there are too many repetitions without explanations within paragraphs [2]. It will be better for you to provide specific details and explanations in your writing [3].</i>	You are really good at using sentences. You don't need to worry about sentence variation.
Writing	Clear thesis. The paragraphs do relate to the thesis pretty clearly. The introduction and conclusion are well written.	The paragraphs have one central idea which is compared with other authors' claims. <i>Transitions are not seen. If they are added, they can complement the paper and make it more understandable [4].</i>	There is flow from one sentence to another. Sentence variation is good, too.
Fabre	The body paragraphs are well-structured and each of them is about one author's opinion, which is very clear. <i>The conclusion is more like your own opinion. There should be more restatements on the different authors' opinions. [2]</i>	Each paragraph is related to one of the authors. <i>But the topic sentence is not very clear [1]</i> . Transitions are used effectively. Enough details are provided.	Sentences are thorough and easy to understand. There is also a good use of sentence variation.
Crocus	It was pretty understandable but <i>it would be better if you shorten it and just say a main point [3]. It also would be better if you mention that the topic is a controversy in the thesis sentence [4]</i> .	Those paragraphs seem to go back to the thesis sentence, but <i>I did not really see the connections between body paragraphs [5]</i> . But I liked the conclusion paragraph, how you summarized the previous paragraphs and reorganized your ideas.	<i>Some grammar issues [1]. It would be better if you avoid starting a sentence with "since" [2]. It would be much easier to understand if you did not start a paragraph with a quote [3]. Don't forget to write a page number after the author's name when you quote it [4]. You need to have more explanations about the quotes you use [5].</i>
Total Count	Global: 4	Paragraph: 5	Sentence and below: 5 T=14

3.6.2.7 Theme 4: Feedback Provided

The quantity and the quality of feedback provided by the improve and the non-improve groups were compared using the coding scheme introduced in Section 3.7.2.1 of this Chapter. The detailed-level categories were collapsed to create categories that could represent the quality of feedback based on the helpfulness to the writer reported in a previous study (Cho et al, 2006). The collapsed categories of feedback compared between the two groups are as below.

- Total number feedback produced measured by idea units
- Percentage of total feedback broken down into praise, criticism, and reflection
- Percentage of criticism feedback addressing macro vs. micro issues (out of total feedback)
- Percentage of all feedback categorized into general vs. specific comments
- Criticism feedback sub-categorized into problem, solution and problem+solution

The data corresponding to each sub-theme listed above was counted for each student in the improve and non-improve groups. Group means were calculated and compared between the two groups.

3.6.2.8 Theme 5: Feedback Incorporation during Revision

For exploring the incorporation of feedback during revision, the following sub-themes were calculated and compared between the improve and the non-improve groups.

- The percentage of criticism feedback incorporated
- The percentage of global criticism feedback incorporated

Incorporation of criticism feedback was determined by tracking and matching the criticism comments given on a draft of essay with all the revisions made between that draft and the subsequent draft. In other words, if a revision(s) addressed a problem/solution comment provided by peer feedback, the criticism feedback comment was coded as being incorporated. Likewise, if a criticism comment was not reflected in the changes made during

the revision, it was coded as not being incorporated. The percentage of criticism feedback incorporated was calculated by dividing the number of criticism comments incorporated by the total criticism comments provided, multiplied by 100.

Incorporation of global criticism feedback was determined by tracking and matching the criticism comments addressing global features given to a draft of writing with all the revisions made between that draft and the subsequent draft. In other words, if a revision(s) addressed the global problem/solution comment, the global criticism was coded as being incorporated. Likewise, if a criticism comment addressing global issues was not reflected in the changes made during the revision, the feedback was coded as not being incorporated. The percentage of global criticism feedback incorporated was calculated by dividing the number of global criticism comments incorporated by the total global criticism comments provided, multiplied by 100.

CHAPTER 4. RESULTS and DISCUSSION

This chapter reports on the results of the analyses conducted on the data collected from the present study in an attempt to address the research questions stated in Chapter 1. The results of analyses are presented under each research question. The three research questions that guided the analyses are as follows:

1. How do peer feedback comments generated by L1 writers compare to those of L2 writers when produced under comparable conditions using a web-based peer review commenting tool SWORD?
2. How do peer feedback comments of L1 and L2 students change/develop over a semester?
3. How do the feedback and revision activities of students whose writing improved the most over the semester compare to those of students whose writing stayed pretty much the same? How do L1 and L2 students compare within these two achievement groups?

4.1 RESEARCH QUESTION 1: PEER FEEDBACK COMMENTS OF L1 AND L2 WRITERS

How do peer feedback comments generated by L1 writers compare to those of L2 writers when produced under comparable conditions using a web-based peer review commenting tool SWORD?

Overview

In contrast to the commonly held belief that L2 students are less productive than L1 students in generating feedback in the target language, little difference was found between the two groups on the overall quantity of feedback produced. In terms of overall participation, L2 students participated at a higher level than their L1 counterparts. When the feedback comments were compared by types, L1 and L2 learners generally produced similar feedback profiles but differed in the proportions of feedback in the ‘problem only’ and ‘solution only’ categories of criticism comments. In addition, compared to L2 learners’, L1 students’ feedback followed a typical feedback structure of praise-criticism-praise and had more

mitigation language when providing negative feedback. The two groups also differed in the construction of solution comments especially with selection of modal verbs. When compared by gender, male and female students produced similar quantities and kinds of feedback.

4.1.1 Students' Level of Participation

In the classes that participated in the current study, students were required to participate in three peer review sessions during each of which they were asked to review and comment on four of their peers' essays. However, their actual participation, number of essays reviewed, and the content of feedback they produced were not reviewed nor graded by the instructors. As a result, while students were highly encouraged to participate, the actual number of essays students reviewed, the volume of feedback comments generated, as well as the number of feedback sessions students participated varied widely.

Of 69 total students enrolled in four classes that participated in the study (L1 = 36, L2 = 33), 66 students gave consent for participation (L1 = 34, L2 = 32). Of the sixty-six students who participated in the study, the number of peer review sessions participated, the number of essays reviewed, and the total number of idea units and words produced by each language group are presented in Table 4.1.1. Because seven students (L1 = 6, L2 = 1) did not participate in any peer review session, those students were not counted when the per capita averages were computed for the number of papers reviewed and the number of ideas and words generated. In other words, the number of papers reviewed and the number of ideas and words produced were divided by $N = 28$ for L1 and $N = 31$ for L2 groups. While counting in the ones who generated zero review would more accurately reflect the level of students' motivation and participation in peer review activity, it would deflate and may distort the per student performance of the L1 group since L1 had higher proportion of students who produced zero comments than L2 (L1 = 17.65%, L2 = 3.13%). Thus, to prevent such distortion, only the students who reviewed at least one essay and generated feedback were counted in calculating the group average.

Table 4.1.1 Students' Participation Level in Peer Review Sessions

	L1	L2
Student-Participants	34	32
Three Peer Review Sessions	12	20
Two Peer Review Sessions	13	8
One Peer Review Session	3	3
Zero Peer Review Session	6	1
Essays Reviewed	249	304
Per Student	8.9	9.8
Idea Units Generated	1,986	2,889
Per Student	70.93	93.2
Words Generated	26,608	32,543
Per Student	950	1,049

As shown in Table 4.1.1, L2 students participated more actively in peer review sessions than L1 students. L2 students participated in more sessions of peer review than L1 students, and reviewed more papers over the semester (L2 = 9.8, L1 = 8.9). While 62.9% of L2 students participated in all three sessions, only 35.3% of L1 students participated in all three sessions. Similarly, the percentage of students who participated in one or fewer sessions stood at 12.5% for the L2 group, whereas the figure was more than twice as high in the L1 group, standing at 26.5%. Based on these figures, it can be argued that L2 students' participation level in peer review was generally higher than that of L1 students.

Further, when the total amount of feedback comments generated is compared between the two groups, L2 learners produced more idea units (L2 = 93.2, L1 = 70.9) and words (L2 = 1,049, L1 = 950) per person during the semester than L1 learners on average. Although the language groups were confounded with instructors in the study, the finding that L2 students produced more feedback comments both in terms of idea units and words is worth noting as it counters the common perception that the limited linguistic capacity and fluency would make L2 students less productive in generating feedback in their target language than L1 students. At least for the L2 students in the current study, who are at intermediate or higher level of proficiency in the target language, it seems that the language proficiency was not an inherent barrier to actively participating and producing feedback comments.

4.1.2 Statistical Test of Differences in Feedback Quantity

In order to explore further whether the differences found in L1 and L2 groups' feedback comments have statistical significance, a MANOVA (multivariate analysis of variance) was conducted on the following four variables: 1) average number of papers students reviewed in a semester, 2) average number of idea units generated per review, 2) average number of words generated per essay, and 3) number of words generated per idea unit.

A MANOVA was conducted that included data from 59 students who participated in peer review at least once using SPSS to see if the two groups' performances show significant difference when all four variables are taken into account. The analysis revealed a significant difference between the two groups when all four variables were considered [$F(1, 57) = 2.74$, $p < .05$]. In the univariate analyses that followed, the number of words per idea unit was found as the only variable showing statistical significance between the two groups [$F(1, 57) = 4.93$, $p < .05$]. The other three variables showed differences in mean that were not at a statistically significant level. The means, standard deviation, and the level of significance from the univariate analyses are provided in 4.1.2.

Table 4.1.2 Comparison of Number of Reviews, Idea Units, and Words

	N	L1 (N=28)		L2 (N=31)		Sig.
		Mean	SD	Mean	SD	<i>p</i>
Essays Reviewed	59	8.89	2.80	9.81	2.90	$p > .05$
Idea Units per Essay	59	7.97	2.00	9.51	4.10	$p > .05$
Words per Essay	59	106.86	41.60	107.05	53.60	$p > .05$
Words per Idea Unit	59	13.35	3.50	11.29	3.60	$p < .05$

Comparison of means show that, although not statistically significant, L2 students reviewed more essays than L1 students throughout the semester (L1 = 8.89, L2 = 9.81), generated more idea units for each essay reviewed (L1 = 7.97, L2 = 9.51), and wrote slightly more words for each review (L1 = 106.9, L2 = 107.0).

However, L1 students used significantly more words for each idea unit than L2 learners ($L1 = 13.35$, $L2 = 11.29$), suggesting that although L1 reviewers commented on fewer topics, they elaborated and provided lengthier comments on each one. This difference could be due to L2 students' limited linguistic capability to be eloquent. Additional explanation for this gap can be found from the attitudes and manners in which the two groups engaged in the peer review activity. A review of students' feedback data showed that while most of L2 students responded to all four feedback prompts provided, L1 students were more likely to respond to only two or three prompts in elaboration.

Generally, L2 students engaged in peer feedback activity with relatively little discretion where they strictly followed each step and procedures with little variation among the students. On the other hand, L1 students allowed room for their own discretion during the peer feedback sessions. For instance, they varied in the number of prompts they responded to, the number of papers they reviewed, and the number of peer feedback sessions they participated. This difference in the two groups' approach to peer review and feedback may be attributable to the differences the two groups have in their past peer feedback experiences and to their overall cultural differences towards school work. A background survey demonstrated that on average L1 students had 3.9 sessions of peer feedback experiences prior to taking the course and resided in the U.S. for more than 10 years, the longest answer choice on the survey. On the other hand, L2 students had two sessions of peer review prior to the class and resided in the U.S. or other English speaking countries for 1.7 years.

4.1.3 Gender Effect

Due to the scarcity of reports on gender differences in feedback studies particularly in the L2 literature, the analyses of the same four variables were conducted to examine gender effect. MANOVA showed that there is no overall gender effect in the feedback produced by male and female students [$F(1,57) = .57, p > .05$]. The univariate analyses that followed also showed no difference between the two genders in any of the four variables and no interaction between gender and language. These results indicate that regardless of writers' language group, male and female students generated a similar number of idea units for each essay they

reviewed ($M = 8.13$, $F = 9.37$), similar number of words for each paper reviewed ($M = 98.97$, $F = 114.18$), and used a similar number of words to explain each idea ($M = 12.29$, $F = 12.24$). The means and standard deviation of male and female students on each of the variables measured are provided in Table 4.1.3.

Table 4.1.3 Gender Effect on Number of Reviews, Idea Units and Words

	N	Men		Women		Sig.
		N=28		F=31		
		Mean	SD	Mean	SD	<i>p</i>
Papers Reviewed	59	9.53	2.60	9.23	3.10	$p > .05$
Idea Units Per Review	59	8.13	2.70	9.37	3.70	$p > .05$
Words Per Review	59	98.97	41.00	114.18	53.00	$p > .05$
Words Per Idea Unit	59	12.29	3.90	12.24	3.60	$p > .05$

4.1.4 Comparison of Feedback Categories

Students' feedback comments were coded into sixteen categories as defined in Chapter 3, Table 3.2. A summary of the feedback categories is provided in Table 4.1.4. For missing data, list-wise deletion was used for more accurate analysis of data, which sets higher standards than may be necessary.

Table 4.1.4 Feedback Categories Used for Coding

Feedback Categories	
1. Praise \pm Specific	
2. Criticism	
A. Meaning-Level Scope	B. Surface-Level Scope
i. Problem \pm Specific	i. Problem \pm Specific
ii. Solution \pm Specific	ii. Solution \pm Specific
iii. Prb+Sol \pm Specific	iii. Prb+Sol \pm Specific
3. Reflection \pm Specific	

4.1.4.1 High-level Categories

When the feedback comments were divided into praise, criticism, and reflection, L1 and L2 groups showed similar profiles across these three categories. For both groups, praise was the most frequent type of comment (L1 = 46.19%, L2 = 46.41%), followed by criticism (L1 = 38.73%, L2 = 38.36%) and reflection (L1 = 14.3%, L2 = 11.6%). These results are in line with earlier studies reporting that students tend to provide substantially more praise than instructors (Cho et al., 2006; Patchan et al., 2009), but the proportions of praise comments from students in the previous studies did not near 50% as they do here. The high proportion of praise may be attributable to the wording of the feedback prompts in the present study, e.g., “Does the paper have a clear thesis?”, where positive responses to the feedback prompts were coded as praise.

L1 and L2 groups did not differ significantly in the distribution of high-level feedback [$F(1,30) = 1.72, p > .05$]. The mean and SD of each group are provided in Table 4.1.5.

Table 4.1.5 Distribution among Praise, Criticism, and Reflection Comments

	N		L1		L2		P
	L1	L2	Mean	SD	Mean	SD	
Praise	12	20	46.19	14.23	46.41	8.63	$p > .05$
Criticism	12	20	38.73	16.65	38.36	11.11	$p > .05$
Reflection	12	20	14.03	4.92	11.60	4.66	$p > .05$

4.1.4.2 Detailed Analysis of Criticism Category

4.1.4.2.1 Function Subcategory: Problem, Solution, Prob+Sol

Even though L1 and L2 students generated similar proportions of criticism comments as a whole, they differed in the sub-types of criticism comments they provided. When the feedback of the two groups is compared at this sub-category level using MANOVA, a statistical significant difference was found [$F(1,30) = 3.15, p = .04$, partial eta square = .25, observed power = .67].

L2 students were significantly more likely than L1 students to provide criticism that only discussed a problem (L1 = 15.0%, L2 = 23.0%, $F(1,30) = 7.95, p = .01$). Conversely,

L1 students were marginally more likely than L2 students to provide comments that only offered solutions (L1 = 13.51%, L2 = 8.37%, $F(1,30) = 3.87$, $p = .058$). They did not differ in the proportion of comments that discussed both a problem and a solution (L1 = 10.22%, L2 = 7.0%). The detailed results of these analyses are provided in Table 4.1.6.

This difference might be due to the varying stages of their linguistic and writing development. Ability to identify solutions to problems in writing is considered a more advanced skill than being able to identify problems. Scardamalia and Bereiter (1983) found that children writers in their study were as good as adult writers in choosing general problem definitions but were not able to define the problems more explicitly nor select a means of dealing with it. L2 students' less practice in writing academic papers in English coupled with less experience in giving writing feedback may account for their lower frequency in providing solutions compared to their L1 counterparts. This finding is consistent with results of a previous study that compared comments of student peers to instructors who were either subject-matter experts or writing experts (Patchen et al., 2009). Writing instructors were found to provide more solutions, content instructors more problems, and students fell in between the two instructors. These differences were explained by the difference in orientation of each group.

Table 4.1.6 Comparison of Sub-categories of Criticism

	N		L1		L2		F	p
	L1	L2	Mean	SD	Mean	SD		
Problem_Only	12	20	15.00	7.82	23.00	7.73	7.95	$p = .01$
Solution_Only	12	20	13.51	8.40	8.38	6.30	3.87	$p = .06$
Problem+Solution	12	20	10.22	6.21	7.00	6.75	1.82	$p = .19$

4.1.4.2.2 Scope Subcategory

When the feedback comments were divided based on the scope of the features the feedback addressed, both groups were found to comment twice or more often on meaning/macro-level features than on surface/micro-level features (see Table 4.1.7). MANOVA showed that the two groups did not differ in terms of the scope of the features

their feedback addressed, [$F(1,30) = 0.50, p > .05$]. These results are counter to earlier studies that report that students in general give more surface/micro-level feedback, and that L2 learners pay more attention to surface-level features than meaning-level features. Two factors seem to have affected this. First, the students participating in the study received a brief training on peer feedback with examples of helpful and not helpful comments with emphasis on the importance of global feedback as part of their classroom lecture. Wallace and Hayes' (1991) found that a training as short as eight minutes aimed at extending students' task schema of revision to include global changes to the text was effective in increasing global revision of students. Second, the format of peer review where students were asked to provide feedback on a separate feedback form outside the essays and the feedback prompts that explicitly asked about high level organization of the text may have directed students to attend more on global issues. Liu and Randall (2004) studied the effect of different modes of feedback and found that students who used MS Word Edit function made twice as many as surface-level feedback compared to students who used peer feedback template with prompts on global features. The present study shows that not only L1 students but also L2 students who are repeatedly found to be skewed towards attending to surface/low-level writing features can be steered toward attending to and comment on meaning/macro-level features of texts at similar percentages to that of L1 students.

Table 4.1.7 Comparison on Meaning-level vs. Surface-level Feedback

	N		L1		L2		F	p
	L1	L2	Mean	SD	Mean	SD		
Meaning-level	12	20	26.36	7.95	27.71	8.04	0.34	$p > .05$
Surface-level	12	20	12.37	10.16	10.65	5.95	0.21	$p > .05$

4.1.4.3 Specificity

Both groups of students made more general feedback comments than specific feedback (See Table 4.1.8). The specificity of the comments did not differ for the two language groups [$F(1,30) = 1.41, p > .05$]. Univariate analyses that followed showed that L1 students gave more specific feedback and less general feedback (L1S = 44.16%, L1G =

53.32%) when compared to L2 students (L2S = 38.25, L2G = 57.20%), but no statistical difference was found between the two groups. No interaction was found between the language group and the specificity of feedback. The detailed figures from the analyses are provided in Table 4.1.8.

Table 4.1.8 Distribution of General vs. Specific Feedback

	N		L1		L2		F	p
	L1	L2	Mean	SD	Mean	SD		
General Feedback	12	20	53.32	11.05	57.20	12.23	1.60	$p>.05$
Specific Feedback	12	20	44.16	10.12	38.25	14.11	0.81	$p>.05$

4.1.5 Comparison of Textual Features of Feedback

Two aspects of the phrasing of the comments: the use of modal verbs and mitigating language were found to show group difference between the L1 and L2 students and were analyzed in depth.

4.1.5.1 Use of Modal Verbs

L1 and L2 students differed in their use of modal verbs. L2 students frequently used obligation/necessity modal verbs such as *should* and *have to* while L1 students employed permission/possibility/ability modals such as *can*, *may*, or *could*.

This difference was particularly evident in solutions. While L1 students rarely used *should*, *must*, or *have to*, these modal verbs were commonly used by L2 students. For L1 students, *could* was the most commonly employed modal verb for offering solutions. Examples of solution comments by L1 and L2 reviewers using distinct modals in expressing somewhat similar ideas are provided in Table 4.1.9.

Table 4.1.9 Comparison of Use of Modals in Solution Comments

L1	L2
For example, you <i>could</i> add a sentence in the end of the introduction stating exactly the two sides of the argument	The paper did not state its thesis at the conclusion so this paper <i>should</i> have a thesis before writing everything.

breaking up big paragraphs into rather smaller ones *could* help in the cohesion of the whole paper.

Paragraphs *should* be condensed in a bigger paragraph.

I think they start the paper very well, but maybe the order of their ideas *could* be changed a little bit to match the opening paragraph more.

The question you asked on the first paragraph *should* be changed, because you mention about bottles being banned but in your paper you never write about it.

L1 students' sentences framed their solutions as examples and possible choices whereas L2 students framed them as logical deduction and/or mandatory actions grounded on commonly accepted presuppositions.

One way to explain this disparity between L1 students' avoidance of and L2 students' preference for obligation/necessity modals comes from changes observed in American and British English in the latter half of the 21st century. After tracing the changes of frequencies of obligation and necessity modals between 1962 and 1992 in the data from the language corpora of British and American English, Smith (2003) reported that in academic prose and other written genres in both dialects, the frequency of "have to" has fallen slightly, the usage of "must" has declined, and the usage of "need to" has overtaken the former two. Smith explained that an increased value placed on equality of power and informality have led to the avoidance of "authoritarian-sounding markers" (p.259). Leech (2003) commented that the decline in the uses of *must* and a shift to *need to* and *should* is possibly "associated with a tendency to suppress or avoid overt claims to power and authority by the speaker or writer," a tendency that "may be called 'democratization'" (p. 237). It is likely that L1 students are more sensitive and up to date with the trend in the usage of modals in English than the L2 students and as a result, predominantly used possibility/ability modals over obligation/necessity modals.

These differences in use of modals by L1 and L2 students can also be explained by the 'face-saving' model (Brown and Levinson, 1978, 1987) of politeness theory. Simply put, face is 'the public self-image that every member wants to claim for himself (Brown and Levinson 1987: 61). There are two aspects to face, positive and negative, and negative face

refers to individual's right to freedom of action and his or her need not to be imposed on by others. Many types of speech acts such as complaints, disagreements and requests intrinsically threaten face, and hence are called face-threatening acts (FTAs). According to Brown and Levinson's model, there are five strategies one can choose from to avoid or weaken an FTA. The more threatening an FTA is, the more polite (higher-numbered) the strategy the speaker must employ to mitigate its effects.

While Brown and Levinson's model provides a good general theoretical and descriptive framework for study of politeness, studies that followed the face-saving theory have indicated that while politeness itself is a universal phenomenon, politeness strategies and individual speech acts may to some extent vary from one language/culture to another. In particular, studies have found that the same speech act may differ in its directness/indirectness in different cultures. In a large scale Cross-Cultural Speech Act Realization Project (CCSARP) and following similar studies it was found that there is an extensive cross-cultural/linguistic variation in directness/indirectness in the expression of speech acts, especially in FTAs such as requests, complaints and apologies.

In addition to linguistic/cultural variations, studies have investigated how a particular type of speech act is performed by non-native speakers. For example, Hinkel (2009) studied the frequency of L1 and L2 learners' use of obligation modals based on the topics of the writing. The study found that the frequency of obligation modals differed significantly depending on the topics of the paper. In essays on politics and racism, L1 students used higher rate of obligation modals than L2 learners, but in topics relating to familiar duties and tradition, L2 learners employed higher rate of obligation modals, and in topics on friendship, L1 and L2 students used a similar rate. This study seems to indicate that students' use of obligation modals is more related to the topic/genre of writing than the level of their target language development. Drawing from this view, the modals used in solution comments by students in the present study may be reflective of students' social and cultural presuppositions about the task of giving feedback on English writing. It appears that L1 students view writing as a heuristic matter, where it is evaluated as better or worse as opposed to right or wrong,

where the purpose of providing feedback is to share reaction as a reader and offer solutions that could help improve the writing. On the other hand, L2 students seem to view English writing from dichotomous standards of right and wrong. Thus their role as a reviewer is to detect the wrongs and offer solutions to make them right. At least for the solutions that they are certain of, they see them as rules that must be adhered to when writing an English paper, not one of choices that are left to the writer to decide on.

Takahashi and Beebe (1993) observed that American English speakers are most likely, and Japanese speakers are least likely to use a positive remark before making a correction, with Japanese speakers of English falling in between. This study suggests that second language acquisition may affect the politeness strategies speakers/writers' employ for various speech acts. Studies seem to suggest that the variations in politeness strategies employed by speakers/writers are embedded in the language they use and second language learners acquire appropriate politeness strategies as part of pragmatic development in language learning.

In light of these findings, it seems reasonable to draw that the differences in use of modals found between the L1 and the L2 groups in the current study is attributed to a combination of factors including social linguistic development in usage of modal verbs in English, cross-cultural/linguistic factor, and the stage of interlanguage development of L2 students. At this point, it is difficult to tease out clearly which factors are more critical on shaping L2 learners' usage of modals or how each factor influenced the use of modals differently.

While some L2 students consistently used obligation modal '*should*', a good portion of L2 students employed both obligation and possibility modal verbs in their solution comments. An examination of their usage of various modal verbs suggests that these students are knowledgeable of the differences between the possibility/ability (*can/could*) and obligation/necessity (*should*) modal verbs, and do possess both types of modals as part of their linguistic repertoire. However, their usage of *can/could* differs from that of L1 students.

The use of possibility/ability modals such as *can*, *may*, *might*, *could* are reported to have a range of textual and pragmatic functions, and together with other linguistic elements, these verbs often serve to mark evidentiality, possibility and likelihood, strategic vagueness, and politeness in discourse (Chafe, 1986; Channell, 1994; Markkanen and Schroeder, 1997; Perkins, 1983). L1 students' use of '*could*' in their solution comments seems to function mostly as a politeness marker with occasional hint of possibility and strategic vagueness. By using '*could*', L1 students are conveying a message that their suggested solution is just one of many possible actions, leaving their peer writers room to reject their solutions. Furthermore, through these politeness markers, they are carefully and actively denoting their intent not to be offensive, authoritative or face-threatening to their peers acknowledging their equal status with the writers.

Conversely, L2 students mostly use *can* to denote possibility and to indicate possible room for improvement. Examples of L2 students' suggestion comments using *should* and *can* are provided in Table 4.1.10.

Table 4.1.10 Examples of L2 students' Use of *Should* and *Can*

	Comments with <i>should</i> or <i>have to</i>	Comments with <i>can</i>, <i>could</i>, or <i>would</i>
Reviewer A	The introduction is too long. You <i>should</i> make it shorter or separate it into two paragraphs.	I think the thesis sentence is the last second sentence in the first paragraph but you <i>can</i> make it more specific and clear like your title.
Reviewer B	You <i>should</i> make the sentence more variable.	It <i>would</i> be better if you use some transition.
Reviewer C	I think the author is good at making long sentences but <i>should</i> avoid making a sentences too long	if I recommend one thing to make the paper stronger, it <i>would</i> be making transition either at the end or beginning of the paragraph

Reviewer C	You <i>should</i> give more information about the effects of climate changes.	Also, you <i>can</i> use more transition words to make your sentences smoother. Most of the sentences are short declarative sentence. You <i>can</i> try to use more kinds of sentences.
Reviewer D	Sentences are not very well formed. You <i>have to</i> work on the sentence variation and improve spelling and grammar.	The paragraphs have a central idea but i think it <i>can</i> be organized better with just one idea for each paragraph.

The divergence in L2 students' usage of *can* from that of L1 students is in line with the findings from previous studies that call for explicit and persistent teaching of usage of modals to L2 learners (Markkanen and Schroeder, 1997; Greenbaum and Quirk, 1990; Leech, 2005; Leech and Svartvik, 2003; Mauranen, 1997).

The teaching of such complex and culturally- bound usage of modals in written discourse does not often take place in classrooms (Hinkel, 2009). This may have left L2 writers less informed and less attentive to the usage of modals in various speech acts, including suggestions and requests employed in giving peer feedback comments. In order to improve effective communication and to avoid misunderstanding arising from the different usages of modals during peer review sessions by students of various language/cultural background, it is advised to introduce the concept of FTAs and the usage of various politeness strategies in L2 writing classrooms to help develop students' pragmatics and genre knowledge on peer feedback.

Alternatively, a number of researchers investigating L2 writing indicate that formulation of texts in L2 require more attention and mental capacity than producing texts in L1 (de Larios et al., 2006; Zimmermann, 2000). It may be that L2 writers are aware of the politeness discourse but their preoccupation with producing the correct forms and meaning keep them short of attending to pragmatics of politeness aspect of writing.

Lastly, it needs to be noted that a large majority of students in the L2 group in the present study were from Chinese and Korean linguistic/cultural origin. As such, the usage of modals and mitigations found in the L2 group in present study may be biased towards

reflecting a particular group of L2 learners. In a L2 group consisting of students of other linguistic/cultural background, the usage of their modals in making suggestions and criticisms may differ from the findings of the present study.

4.1.5.2 Use of Mitigation Language

For most L1 students, feedback comments followed a positive-negative-positive structure. Table 4.1.11 shows examples of L1 students' mitigation comments inserted before/after the criticism comments. Particularly prior to making a negative comment, L1 students added some sort of mitigation that either expressed empathy to the writer about what the writer had done, included an irrelevant or contradicting praise, or downplayed their criticism. This pattern is consistent with a previous study (Smith, 1997) that reports that teachers' feedback comments usually begin with positive evaluation, move to negative evaluation and coaching and end with coaching or positive evaluation. According to Smith, teachers may choose to provide positive comments before negative in order not to disappoint students, to demonstrate their fairness that they were not only searching for papers' faults, or to motivate students. While these intentions may be good, Smith expressed concern that adherence to a ritual of opening a feedback with positive may diminish the effect of praise because students who recognize positive opening as a generic rule of feedback genre may ignore the meaning of it simply because they appear at the beginning.

Table 4.1.11 Examples of L1 Students' Feedback with Mitigation

Types of mitigation	Feedback examples (mitigation in bold, italic)
Meaningless praise	<p><i>You have really good thoughts</i> but I would suggest to work a little on organizing. In your introduction I would try to be a little more specific with your thesis and what the controversy is between.</p> <p><i>Good main idea</i> but I noticed that your first paragraph talks more about the benefits of electronic media instead of the environmental benefits.</p>
Conflicting message	<p><i>I like the background information in the introduction</i> but I think that it is too long compared to the actual rhetorical analysis of your paper.</p>

<p>Empathize with writer</p> <hr/>	<p>As for sentence flow, there were a few places where you switched topics abruptly, <i>but they should be fine.</i></p> <p><i>The thesis is pretty clear</i> although it seems to be split up over multiple sentences which can seem, at least to me, a little confusing upon first read.</p>
<p>Downplaying criticism</p> <hr/>	<p><i>I know this is just a rough draft so I don't want to give you too much criticism over it</i> but you have a few punctuation errors that you should be able to catch if you reread your paper.</p> <p><i>I understand what the controversy is about</i>, but the thesis in the paper needs more explanation and clarification.</p> <p>I did see typo almost at the end of page 3 there are 2 together, <i>although is practically not important.</i></p> <p><i>I am not a good person at grammar</i>, but what I would highly recommend you is to go to the W Center and ask someone to revise it.</p> <hr/>

In contrast, the positive-negative-positive structure was not prominent in L2 students' feedback in general and little mitigation was found in L2 students' feedback. In cases where there were some mitigation, majority of them were downplaying comments stemming from their status as L2 learners. Unlike L1 students who are relatively acquainted with the generic structure of feedback genre, L2 students seem to react differently to this 'ritual' praise. A study that explored L2 learners' acceptance and reaction to native teachers' written feedback, (Hyland & Hyland, 2001) reports that some L2 learners were confused by and misunderstood some of their teachers' feedback comments. L2 students with less exposure to and experience with the genre of feedback may give these praises equal significance as the negative comments that follow and experience confusion and misunderstanding.

Another noticeable characteristic of L2 students' feedback was that they were forthright in stating their criticism. This may be due to the lack of linguistic competency to produce elegant and sophisticated prose, as was noted as a key characteristic of L2 learners' writing (Hinkel, 2003), or due to unfamiliarity with the genre of English peer feedback that includes large amounts of praise and mitigation comments (Patchen et al., 2009). In general, L2 learners used few to no qualifiers in their criticism. This directness, however, may create a feeling of harshness and be viewed as offensive by the writers.

Table 4.1.12 Examples of L2 Students' Criticism Feedback without Mitigation

Degree of Mitigation	Examples of L2 Students' Feedback
Minor Mitigation	There are grammar mistakes throughout the essay. At first it didn't really matter but when they started showing up all the time it took my focus away from the content of the paper. Work on that. Also try to vary more with word choice still, for example different verbs.
No Mitigation	Your topic sentences are not clear. And there are too many repetitions without explanations within paragraphs.
No Mitigation	The paragraphs and the ideas in this essay are not organized.
Minor Mitigation	There are transitions but they are not very effective, it's hard to make connections of the three body paragraphs.

4.2 RESEARCH QUESTION 2: LONGITUDINAL CHANGES IN COMMENTS

How do peer feedback comments of L1 and L2 students change/develop over a semester?

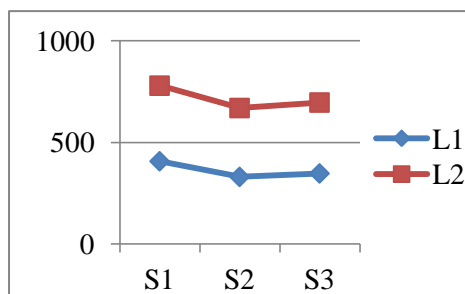
In order to investigate how students' feedback comments changed/evolved over a semester across three feedback sessions and to explore whether the changes within each group converged or diverged, the analyses were conducted in two steps. First, the quantity of feedback units and words generated for each session were compared across sessions to examine the changes in production over time. Second, the distribution among various types (categories) of feedback comments were computed for each session and compared across the three sessions to understand how the characteristics of the feedback comments changed over time. The changes within each group were compared against the other group to investigate whether the two groups' feedback comments evolved in a similar or opposite direction. A within-subject repeated measure design was employed with time and language groups as the main factors of analyses. Since the main interest of the analyses was to see longitudinal development in feedback comments, all missing data were treated as list-wise deletion and only the feedback comments of the students who participated in all three sessions were included in the analyses. Out of total 66 participants, only 32 students participated in all three sessions (L1 = 12, L2 = 20).

4.2.1 Quantitative Developments

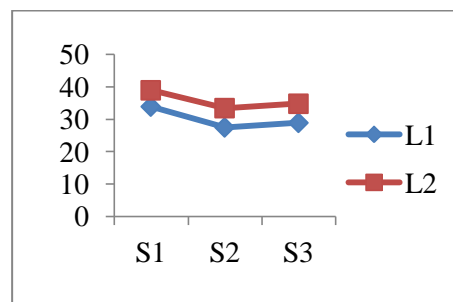
Overall, both groups produced less feedback over time; students generated fewer idea units per review and became less verbose on each idea unit over the semester. (See Figure 4.1.) Generally, for both groups, the degree of change in the feedback was greater between S1 to S2 than between S2 to S3.

As shown in Figure 4.1a, L1 students generated 18.92% fewer idea units between S1 to S2 and 5.15% more idea units from S2 to S3 ($407 > 330 > 347$), and L2 group's idea units decreased by 14.12% from S1 to S2 and increased 3.97% from S2 to S3 ($779 > 669 > 696$). As shown in Figure 4.1 (d), the L2 words per idea unit decreased more sharply than the L1 group. L1 students used 17.98% fewer words per idea unit between S1 to S2 and 5.88% more words per idea unit from S2 to S3 ($16.80 > 13.78 > 14.59$). L2 students used 21.3% fewer words per idea unit from S1 to S2 and 23.08% fewer words from S2 to S3 ($14.92 > 11.74 > 9.03$) widening the gap between the two groups over time.

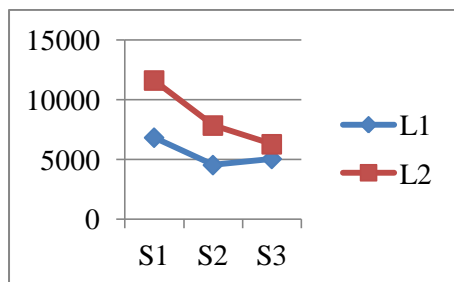
Figure 4.1 Changes in Number of Feedback Ideas and Words across Sessions



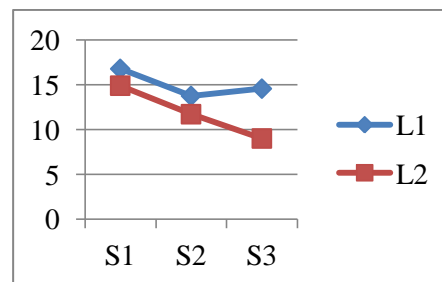
(a) Changes in Total Idea Units



(b) Changes in Idea Unit per Student



(c) Changes in Total Words



(d) Changes in Word per Idea Unit

These data simply demonstrate that the students generated less feedback as the feedback sessions were repeated. These data do not provide any explanation about how the feedback comments changed over time: whether the feedback became more helpful or not. To address this issue, changes in the types of feedback were compared across sessions.

4.2.2 Changes in Feedback Categories

The changes in the distribution of feedback categories, which reflect the quality/helpfulness of feedback, were compared across sessions within a group to examine if and how the feedback comments changed over time. The changes were compared to determine whether the two groups' feedback comments evolved in a similar or opposite direction. A within-subject repeated measure design was employed with time and language groups as the main factors of analyses. The changes in the feedback categories were investigated in three dimensions: 1) distribution of feedback in high-level categories of praise, criticism, and reflection; 2) distribution of feedback in the sub-categories of criticism by function (problem and/or solution); scope (meaning-level or surface-level); and 3) distribution of feedback in specificity (general vs. specific).

The results of analyses show that there was no statistically significant main effect of time across the three feedback sessions in any of three dimensions. Main effect of group was found significant only in the sub-categories of criticism by function. The interaction between the time and groups was not found statistically significant in any analyses.

These suggest that overall, the types of feedback comments generated by L1 and L2 groups were not significantly different from each other except for the criticism subcategories of '*problem-only*', '*solution-only*', and '*problem and solution*'; did not show significant change in the distribution of feedback types over time; and whatever changes that occurred in the distribution of feedback in each group did not differ significantly between the two groups.

Nonetheless, analyses on a few measures reached close to statistically significant levels in the main effect of time and in the interaction between time and group. This implies that although the changes measured are not big enough to demonstrate statistical significance, the students' feedback comments do display some changes over time and the two groups do

evolve differently in certain categories of feedback. The results of each analysis are provided in detail under each section.

4.2.2.1 High-level Categories

The distribution of feedback comments at high-level categories produced no significant main effects for time [$F(1,30) = 2.30, p = .07$] or group [$F(1,30) = 1.72, p = .19$]. Nor was the interaction of the two significant [$F(1,30) = 0.97, p = .47$]

The two groups showed little difference in the feedback profile they generated at high-level categories of praise, criticism and reflection. The ratio between praise and criticism comments was more or less the same for both L1 and L2 groups. However, although statistically not significant, both groups gave more praise comments (S1 = 36.69%, S2 = 47.59%, S3 = 51.71%) and less criticism feedback (S1 = 43.48%, S2 = 36.67%, S3 = 35.36%) as time passed. The percentage of reflection comments went down across sessions in both groups (S1 = 13.89%, S2 = 11.85%, S3 = 11.80%). The mean percentage of feedback comments for each category per group per session is provided in Table 4.2.1.

Table 4.2.1 Mean Percentage Changes in Praise, Criticism, Reflection over Time

	L1 Group			L2 Group		
	Session 1	Session 2	Session 3	Session 1	Session 2	Session 3
Praise	42.27	49.07	47.23	38.14	46.71	54.40
Criticism	41.03	36.38	38.78	44.94	36.84	33.30
Reflection	15.16	14.31	12.60	13.11	10.38	11.32

The lack of improvement over time is somewhat surprising. In the field of education in general and in writing instruction, repeated exposure and practice usually brings enhancement in performance (Kellogg et al., 2009). Under this view, the proportion of criticism comments were expected to increase and that of praise comments to decrease over time as students perceive criticism comments more helpful than praise (Nelson et al., 2009). Also, the repeated experience of reviewing peers' paper was expected to enhance students' critical reading and self-reflection capability and thus increase the portion of reflection

comments. The changes that took place, albeit not significant, were in the opposite direction to the assumption where the praise comments increased and criticism and reflection comments declined.

One may argue that the improvement in students' writing achieved over three assignments may be responsible for the increase of praise comments. Although the students' writing quality is reported to have improved over the semester according to the scores given by the instructors, a closer examination of the praise comments show that the increase came mostly from the growth in general praise comments (S1 = 29.4%, S2 = 38.6%, S3 = 40.4%) and not so much from specific praise comments (S1 = 10.3%, S2 = 9.3%, S3 = 11.4%). In most cases, general praise comments were feedback that rubber stamped the paper by responding positively, such as commenting 'yes', to the feedback prompts. These figures together with quantitative analyses that show reduced feedback over time seem to indicate that students became less enthusiastic about peer feedback and their performance on peer feedback deteriorated in quantity and quality over time.

This counter-intuitive result could be due to the fact that the students were not recognized/rewarded or enforced to provide more helpful comments. In addition, the same feedback prompts used in all three feedback sessions may have diminished the novelty effect and reduced students' interest towards peer feedback over time. According to Ericsson (2006), developing expertise in physical or cognitive skills require four conditions: a) effortful exertion to improve performance, b) intrinsic motivation to engage in the task, c) carefully tailored practice tasks, d) feedback that provides knowledge of results, and e) high levels of repetition. Students in the present study had conditions c), d), and e) fulfilled to a certain degree but conditions a) and b) were not satisfied. It can be posited that without proper reward or control for the task, students do not exert efforts required to improve performance and their motivation to engage in the task is not sustained over time when students engage in the same routine task. It seems that the students came to provide feedback that were quicker and easier to give and shifted their efforts, attention and time to the tasks that were more recognized, rewarded and/or enforced. While the course instructors in the

present study did not choose to reward students for providing more helpful feedback, students in other SWoRD studies (Cho et al., 2006; Cho & Schunn, 2007) were rewarded for higher back-evaluation ratings received from the writers as part of their grading system. It would be interesting and worthwhile to examine changes over time when the students are rewarded or controlled (through penalty) for the helpfulness of their feedback comments.

The L1 group's third session feedback supports this view. The data show that the quality of L1 group's feedback, manifested by the proportion of praise vs. criticism and general vs. specific comments, generally declined from S1 to S2 but rebounded in S3. On the other hand, the L2 group's performance continuously fell from S1 to S2 to S3. The difference can be attributed to the fact that the L2 instructor gave no feedback or control over the types of feedback the students generated throughout all sessions but the L1 instructor had a 'serious talking to' to the students right before their third session, triggered by some mischievous feedback given by a L1 student during S2.

During administration of the survey, a student voluntarily commented that he had missed some sessions of peer feedback because it was not counted towards the course grade. He said that had it been part of the course grade, he would have participated in peer feedback with greater eagerness.

4.2.2.2 In-depth Analyses of Criticism Category

In the current study, L1 and L2 students differed significantly on whether their criticism comments included a description of both a problem and a solution. As such, longitudinal analyses were conducted to investigate whether changes occurred in this critical category over time.

4.2.2.2.1 Criticism by Function: Problem, Solution, Problem+Solution

Repeated-measure analyses showed a statistically significant main effect of group [$F(1,30) = 3.15, p = .04$] in the distribution of 'problem only', 'solution-only', and 'problem+solution' categories. But the effect of time [$F(1,30) = 1.0, p = .45$], and interaction between time and group [$F(1,30) = 2.36, p = .06$] were found insignificant.

These results indicate that the two groups were significantly different in the overall proportion of feedback among ‘problem only’, ‘solution only’, and ‘problem and solution’ comments as was discussed in Section 4.1.4.2.1—means and standard deviations of each group for the sub-categories are provided in Table 4.1.6—but neither group changed significantly in the proportion of feedback types generated in these subcategories over the semester and the two groups did not significantly diverge in the changes they demonstrated across three sessions.

However, it is worth noting that the effect of interaction between time and group was found to be close to significant [$F(1,30) = 2.36, p = .06$] suggesting that although not statistically significant, the two groups did demonstrate some changes over time in how they gave criticism comments and that the two groups changed in different directions on these sub-categories of criticism feedback. L1 group provided more ‘problem only’ criticisms and less ‘solution only’ comments as time passed. Conversely, L2 group gave less ‘problem only’ and more ‘solution only’ comments over time. L1 group’s ‘problem+solution’ comments stayed more or less the same over time whereas that of the L2 group dropped drastically across sessions.

For ‘problem only’ and ‘solution only’ categories, the two groups started off with a fairly wide gap in S1 but converged over time to close the gap across sessions. On the other hand, the two groups started off similarly with the ‘problem and solution’ category but widened the gap over time. It is interesting that the two groups showed reverse trends on ‘problem only’ and ‘solution only’ categories, both groups moving towards a more balanced distribution of feedback profiles.

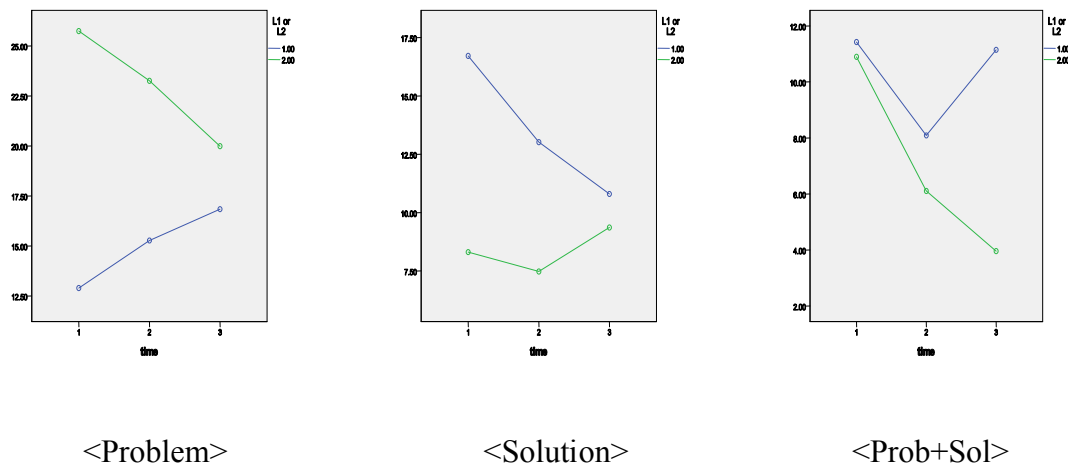
Although exposing students to repeated practice of peer review and feedback did not bring significant changes to the distribution of feedback types over time, it may have helped students shift their orientation to the areas they have been less attentive to previously, bringing both groups of students to have somewhat balanced orientation for both the problems and the solutions.

The mean percentages of feedback in ‘problem only’, ‘solution only’, and ‘problem and solution’ categories per session per group is provided in Table 4.2.3 and graphed in Figure 4.2.

Table 4.2.3 Changes in Problem, Solution, Problem+Solution over Semester

	L1 Group			L2 Group		
	Session 1	Session 2	Session 3	Session 1	Session 2	Session 3
Problem only	12.90	15.27	16.84	25.74	23.25	19.98
Solution only	16.71	13.02	10.80	8.31	7.48	9.36
Problem + Solution	11.42	8.09	11.15	10.90	6.16	3.96

Figure 4.2 Changes in Problem, Solution, Problem+Solution over Semester



4.2.2.2.2 Criticism by Scope: Macro/Meaning-level vs. Micro/Surface-level

When the ratio between the meaning versus surface-level criticism comments were compared across sessions between the two groups, no main effect of time [$F(1,30) = .76, p > .05$], group [$F(1,30) = 1.23, p > .05$], or the interaction between time and group [$F(1,30) = 1.34, p > .05$] were found significant.

Although not statistically significant, both groups generated fewer meaning-level comments over time. As for surface-level comments, L1 group generated more and L2 group

produced less across sessions. The mean percentages of feedback in ‘meaning-level’ and ‘surface-level’ categories per session per group is provided in Table 4.2.4.

Table 4.2.4 Changes in Meaning-level versus Surface-level Feedback over Semester

	L1 Group			L2 Group		
	Session 1	Session 2	Session 3	Session 1	Session 2	Session 3
Meaning-level	29.78	23.42	25.89	30.68	27.57	24.87
Surface-level	13.01	15.07	15.05	15.87	10.19	9.06

4.2.2.3 Specificity: General vs. Specific Feedback

When the ratio between general versus specific comments were compared across sessions between the two groups, no main effects were found for time [$F(1,30) = 1.86, p > .05$], group [$F(1,30) = 1.41, p > .05$], or the interaction between time and group [$F(1,30) = 1.80, p > .05$].

Although not significant, both groups generated more general comments and fewer specific comments across sessions. Both groups moved in the same directions but L2 group showed sharper changes in both categories. The mean percentages of feedback in ‘meaning-level’ and ‘surface-level’ categories per session per group is provided in Table 4.2.5.

Table 4.2.5 Changes in General vs. Specific Feedback

	L1 Group			L2 Group		
	Session 1	Session 2	Session 3	Session 1	Session 2	Session 3
General	49.89	53.55	56.52	48.39	60.58	62.63
Specific	46.92	45.54	40.01	47.29	32.01	35.44

4.2.3 End of Semester Student Survey

The post-semester student survey consists of three sections: Peer Review Experience, Writing Experience, and the SWORD Experience. The Peer Review Experience section of the survey consists of total 41 items but in this section, only the results of 16 items that are

related to longitudinal aspect of peer review and feedback are presented. The complete survey form with all items is provided in Appendix B.

The sample size of this survey is not large ($N = 38$, $L1 = 23$, $L2 = 15$) but it provides voices of the first-hand providers and recipients of the feedback who used peer feedback in their revisions multiple times over a semester. Student survey on the longitudinal effect of peer feedback is rare in the literature. In addition, with its reasonably high face validity and high internal reliability, the survey data serves as an alternate angle to triangulate the findings from the earlier analyses of the study.

The survey analysis included not only the students who generated feedback in all three sessions but also the ones who gave feedback in two sessions and received feedback in three sessions. They were included because otherwise the low response rate to the survey in L2 group would have left the sample size very small at $N = 23$ ($L1 = 12$, $L2 = 11$). The number of students who generated feedback in all three sessions was less than 50% ($n = 32$) of the total participants ($N = 66$). Not surprisingly, the three-session students responded slightly more favorably to the peer review and feedback experience than the two-session students. Although the difference between the two groups was not stark, including only the three-session students' responses in the analysis could result in a bias and may not be representative of general students' reaction to peer review. After all, the students who generated feedback in all three sessions are the ones who are more enthusiastic about peer feedback than the average students. That being said, the students who generated feedback in one or fewer sessions were excluded from the analysis as they lack the mere experience of generating feedback multiple times over a period.

4.2.3.1 Changes in Attitude and Perception

Five items on the survey asked about changes in the students' attitudes and perception towards peer review and feedback (see Table 4.2.6). Overall, for both L1 and L2 groups, a majority of students reported that they became more interested in providing and reviewing peer feedback and that both the feedback they provided and received became more helpful in their revision over time. L2 students reported developing more positive attitude and

perception toward peer feedback over time than L1 students. L1 students reported growing interest and engagement in both areas but more so in reviewing the feedback they received than in providing feedback to peers.

Table 4.2.6 Mean Responses on a Student Survey on Behavior and Attitude towards Peer Feedback on a Six-Point Likert Scale with 1 = Strongly Disagree and 6 = Strongly Agree

Item	Question	L1 (n=23)		L2 (n=15)	
		Mean	SD	Mean	SD
2.1	Over the semester, I became more interested in seeing my peers' feedback on my paper.	3.96	1.09	4.21	1.05
3.1	Over the semester, I became more interested and engaged in providing feedback.	3.96	1.14	4.00	0.59
2.4	Over the semester, the feedback I received became more helpful in my revision and writing.	3.96	1.20	4.29	0.91
3.6	Over the semester, I think my feedback became more helpful to peers in their revision and writing.	4.00	1.19	4.07	0.99

4.2.3.2 Perceived Changes in Incorporation of Feedback

The survey asked five questions about the degree of incorporation of the feedback they received. When asked whether they incorporated more feedback in revision over time, a majority of students in both groups (L1 = 78.2%, L2 = 86.7%) responded that they did (see Table 4.2.7). When asked to indicate the percentage of feedback they incorporated in their revision for each session (open ended question), students in both groups, on average, reported higher percentage of implementation across sessions. The average incorporation rates for each session reported by students were 32.4%, 39.8%, and 47% for L1 group and 43.3%, 47.7%, and 48.0% for L2 group.

Contrary to previous studies that claim L2 students' low incorporation of and low preference for peer feedback due to the distrust L2 students have towards their peers (Carson & Nelson, 1996; Zhang, 1995), the present survey shows that L2 learners' acceptance rate of

peer feedback is, in fact, considerably higher compared with that of their L1 counterparts, especially during the first two sessions. L1 students showed strong distrust for the peer feedback they received in S1 with 82.6% (vs. L2 = 53.3%) of the total respondents incorporating 50% or less of the peer feedback. Even after the steady increase of the rate over time, only 30.35% (vs. L2 = 60%) of the students accepted more than 50% of the peer feedback in S3.

These data imply that the widely-held perception in the L2 community that L2 learners distrust the feedback comments from their peers mainly due to the peers' limited language proficiency may be misleading. L2 students, at least when based on self-report, demonstrated stronger trust towards the feedback from peers compared to L1 group with similar age and educational background, through higher percentage of peer feedback incorporation during their revision.

Moreover, the wide gap between the L1 and L2 groups in the feedback incorporation suggests that the limited language capacity of L2 learners may be affecting the incorporation level of peer feedback in the opposite way than what has been posited from earlier studies. L2 students who are less confident and less certain about their writing in target language may be in need of, and thus rely more on the feedback they receive for revision. This may affect L2 students to be less critical about the feedback from peers than the L1 students.

Table 4.2.7 Percentage of Reported Feedback Incorporation by Group and by Session

Item	Question	L1 (n=23)	L2 (n=15)
2.2	Over the semester, I accepted/incorporated more peer feedback comments in my revision.	M=4.35, SD=1.07	M=4.25, SD=1.13
5.1	Please indicate the approximate percentage of peer feedback you incorporated in your revision from S1.	M=32.4% (0~50%=82.6%, 51~100%=17.4%)	M=43.3% (0~50%=53.3%, 51~100%=46.6%)
5.2	Please indicate the approximate percentage of peer feedback you incorporated in your revision from S2.	M=39.8% (0~50%=73.9%, 51~100=21.7%)	M=47.7% (0~50%=46.6%, 51~100%=53.3%)

5.3	Please indicate the approximate percentage of peer feedback you incorporated in your revision from S3.	M=47.0% (0~50=69.6%, 51~100=30.4%)	M=48.0% (0~50%=40%, 51~100%=60%)
-----	--	---	---

When asked why their level of incorporation had increased or decreased, L1 students cited ‘better’ feedback as the biggest reason (55%), followed by ‘worse’ feedback (25%) and ‘other factors’ (20%). (See Table 4.2.8) L2 students gave ‘other factors’ as the biggest reason (54.5%), followed by better feedback (27.3%) and worse feedback (18.2%). These results seem to be in line with the findings from the analyses of feedback categories in Section 4.1. In other words, the sheer quantity and the quality of feedback comments did not increase over time, but students reportedly became better at making sense of and utilizing the feedback they received in their revision over time.

Table 4.2.8 Reasons for Change in Feedback Incorporation over Time

Reason	Description	L1 (n=23)		L2 (n=15)	
		person	%	person	%
1. Better Feedback	More Specific, Detailed, Helpful, Accurate Feedback	11	55.0%	3	27.3%
2. Worse Feedback	Less Appropriate, Helpful, Serious Feedback	5	25.0%	2	18.2%
3. Other Factors	More familiarity with (feedback/revision)process, learned to do peer review better, did more revision/better revision	4	20.0%	6	54.5%
4. N/A	No response	3	13.0%	4	25.7%

After investigating the various features of feedback and the factors affecting implementation of feedback, Nelson and Schunn (2008) concluded that understanding is the only significant mediator that affects implementation of feedback. The students in the present study reported implementing more feedback over time even though the quality of feedback did not improve. Thus, it can be posited that the students’ ability to understand and make sense out of the feedback they received improved over time and that they have become more skillful feedback users. L2 students may have achieved greater gains in learning to better

utilize peer feedback because they have been less exposed to peer feedback activity in the past than the L1 group.

4.2.3.3 Perceived Changes in Feedback Skills

While the feedback output seems not to have improved significantly over time, the findings from earlier sections seem to suggest that students have become better at using peer feedback over time. To investigate this aspect further, five questions relating to students' perception on the changes to feedback skills were analyzed. Most students (L1 = 87%, L2 = 86.7%) in both groups reported that they came to better understand/make sense out of the feedback comments they received over time. (See Table 4.2.9) More students in the L2 group (86.7%) thought the feedback they received became more helpful in their revision and writing than L1 group (69.75%). Almost all students in both groups (L1 = 95.7, L2 = 93.3%) reported that it became easier to identify problems in peers' paper and most students (L1 = 86.9%, L2 = 86.7%) agreed that it became easier to come up with solutions to the problems they identified and that their feedback became more accurate (L1 = 86.9%, L2 = 86.7%).

These responses seem to indicate that over time, most students experienced improvement in their peer feedback performance. Although the statistical analyses showed that the distribution of feedback types did not change significantly across sessions, the survey shows that the students came to perform various activities pertaining to peer feedback, i.e., identifying problems, devising solutions, understand/making sense out of feedback and applying the feedback in revision, with greater easiness and accuracy.

Table 4.2.9 Students' Perceived Improvement in Peer Feedback Performance

Item	Question	L1 (n=23)		L2 (n=15)	
		Mean	SD	Mean	SD
2.3	Over the semester, I came to make better sense out of/understand better the feedback I received.	4.45	0.86	4.13	1.06
2.4	Over the semester, the feedback I received became more helpful in my revision and writing	3.82	1.14	4.4	0.91

3.3	Over the semester, it became easier to identify problems in my peers' paper.	4.68	0.65	4.73	0.80
3.4	Over the semester, it became easier to come up with solutions to the problems I identified.	4.35	0.85	4.40	0.74
3.5	Over the semester, I think my feedback became more accurate	4.73	0.98	4.20	0.68

4.2.3.4 Perceptions of Importance of Grading Peer Feedback

Since exclusion of peer feedback from the course grading system is one of the plausible explanations for students' decreased interest and engagement in peer feedback and for the lack of an improvement in the quality of feedback over time, students were asked in the survey if they would have behaved differently had peer feedback been part of the course grading system. To this question, 60.9% of L1 students and 46.7% of L2 students responded 'yes'. These figures seem not high enough to posit that including feedback as part of the course grading system would bring changes in students' behavior and the quality of feedback over time. However, the students whose survey responses were analyzed were the students who produced feedback at least in two sessions, who actively participated in peer feedback activity despite the minimal reward it had on the overall grade. When the responses of students with lower participation, who produced feedback in one or less session are looked at, much higher rate of students answered 'yes' (83.3%) on the item. The detailed percentages of responses are provided in Table 4.2.10.

Findings from the survey seem to suggest that including peer feedback as part of course grade could bring more active and sustained interest and participation of students across the semester, and possibly result in improved feedback quality over time. In addition, regardless of the quality of feedback students receive, the repeated participation in peer feedback over time seems to help build students' competency in identifying problems, devising solutions, better understanding feedback, and thereby making better revisions. Considering that the capacity to make good revisions are key to becoming good writers, the longitudinal effect of peer feedback seem to be directly related to bringing up good writers.

Table 4.2.10 If Peer Review/Feedback is Graded...

Item	Question	L1 (n=23)	L2 (n=15)
3.9	If your peer review comments were reviewed and graded as part of the course grade, would you have done it differently?	Yes=60.9% No=34.8%	Yes=46.7%, No=53.3%
	If YES to 3.9, what would be the major difference?	4.35%	20.00%
	1.I would have read my peers' paper more thoroughly to provide accurate feedback	43.50%	33.30%
3.9-1	2.I would have tried to provide more detailed and specific comments	30.40%	0%
	3.I would have been more sensitive to deadlines and participate in all sessions=30.4%	4.35%	0%
	4.Other=4.35%		
	*: multiple responses allowed		

4.3 RESEARCH QUESTION 3: IMPROVE VS. NON-IMPROVE GROUPS

How do the feedback and revision activities of students whose writing improved over the semester compare to those of students whose writing did not improve? How do L1 and L2 students compare within these two achievement groups?

In order to investigate the relationship between students' peer review activities and the gains in their writing ability, two groups of sample cases of students were selected: a group of students whose writing score improved over the semester— 'improve group', and a group of students whose writing score remained unchanged or declined over the semester— 'non-improve group'. A summary of the background information of the sample participants is re-presented in Table 4.3.1.

Table 4.3.1 Background and Writing Scores of Sample Cases

[L1 Group]

no	grp	name	L1	yrs in US	sex	age	major	SAT (writing)	writing score
----	-----	------	----	-----------	-----	-----	-------	---------------	---------------

1	High	Ethan	Eng	> 10	M	19	science/engr	640 (12)	F/B-/A-
2	High	Doug	Eng	> 10	M	n/a	Science/engr	520 (8)	B+/B+/A
3	High	Peter	Eng	> 10	M	20	humanities	600 (10)	D+/C-/B+
4	High	Jordon	Eng	> 10	M	18	business	550 (10)	C-/C/B
5	Low	Wendy	Eng	> 10	F	18	undeclared	680 (10)	B-/C+/B-
6	Low	Amy	Eng	> 10	F	18	social science	570 (8)	B+/B+/B+

[L2 Group]

no	grp	name	L1	yrs in US	sex	age	major	TOEFL/ SAT (writing)	writing score
1	high	Megan	Chinese	<1	F	20	science/egnr	98 (25)	92/95/94
2	high	Harry	Finnish	<1	M	19	undeclared	110 (29)	85/94/94
3	high	Steve	Tamil	<1	M	19	undeclared	98 (28)	92/94/95
4	high	Sally	Korean	<1	F	23	humanities	89 (21)	90/91/93
5	low	Yves	Chinese	1-3	F	20	undeclared	90 (*)	86/78/75
6	low	Jennie	Korean	4-7	F	19	science/egnr	97 (22)	94/93/90
7	low	Heather	Chinese	4-7	F	21	science/egnr	580 (9)	94/89/90

* : not reported

The ways in which the sample students engaged in writing, peer review/feedback, and revision were examined in-depth around the five themes and their sub-themes identified in Chapter 3 as the factors affecting students' writing development over a semester. These themes and sub-themes are re-presented in Table 4.3.2.

Table 4.3.2 Themes and Sub-themes

Theme	Sub-themes
1. Types of Revisions	<ul style="list-style-type: none"> a) Changes in surface features such as punctuation and grammar b) Changes in style/expression c) Extension/reduction of existing idea (in part) d) Addition/deletion of new idea (in whole) e) Change in organization: relocation/shifting of texts across sentence or paragraph boundaries
2. Number of Revisions	

Changes that constitute a single type of revision were counted as one revision regardless of the amount and/or the magnitude of changes made to the text.

3. Feedback Received

- a) Number of criticism comments received (in idea unit)
- b) Number of global-level criticism feedback received

4. Feedback Provided

- a) Total number of feedback comments generated (in idea unit)
- b) Proportions of praise, criticism, and reflection comments
- c) Percentage of macro vs. micro comments
- d) Percentage of general vs. specific comments
- e) Percentage of problem vs. solution vs. problem+solution comments

5. Incorporation

- a) Percentage of criticism feedback incorporated
- b) Percentage of global criticism feedback incorporated

An in-depth analysis of students' behavior and performance on the themes revealed that the 'improve' and the 'non-improve' groups behaved differently with regards to some themes and similarly on others. Following the approach adopted by many expert vs. novice studies, the conditions that are associated with the improve group are viewed as variables fostering growth in students' writing ability whereas conditions associated with the non-improve group are considered inhibitors to development of writing competency. Some themes were found to have stronger relationship with the students' writing development than others. The types of revision strategies students employed were found to be critical in differentiating the 'improve' and the 'non-improve' groups. The percentage of incorporation of criticism feedback in general and global-level criticism comments also divided the 'improve' and the 'non-improve' groups. On the other hand, the types of peer feedback comments generated by the students did not differ between the 'improve' and the 'non-improve' groups. A longitudinal analysis suggested that the way students engage in revision not only have an immediate effect on the quality of the subsequent draft of the essay but may have a long-term effect on the following assignments. A brief summary of the topics and findings that will be presented in the following sections is provided in Table 4.3.3.

Table 4.3.3 Summary of Topics and Findings

Section	Topic	Findings (Improve vs. NonI Grps)	L1 vs. L2
4.3.1	Revision		
4.3.1.1	Revision Repertoire	I = NI	L1=L2
4.3.1.2	Amount of Revision	I > NI (L2), I = NI (L1)	L1≠L2
4.3.1.3	Types of Revision: Type 4	I > NI	L1=L2
4.3.2	Incorporation of Feedback		
4.3.2.1	Feedback Received		L1≠L2
4.3.2.1.1	Criticism Feedback	I > NI (L1), I = NI (L2)	L1≠L2
4.3.2.1.2	Global-level Criticism Feedback	I > NI (L1), I < NI (L2)	L1≠L2
4.3.2.2	Feedback Incorporation		
4.3.2.2.1	Incorporation of Criticism Feedback	I > NI	L1= L2
4.3.2.2.2	Incorporation of Global Criticism Feedback	I > NI	L1= L2
4.3.3	Feedback Provided	I = NI	L1= L2
4.3.4	Longitudinal Effect of Feedback and Revision		L1= L2

4.3.1 Revision

4.3.1.1 Revision Repertoire

The breadth of revision types/strategies utilized during revision was compared between the ‘improve’ and the ‘non-improve’ groups to examine its effect on the quality of students’ revision and the subsequent paper. The five types of revision strategies require distinctive skills for implementation. Type 1 (Punctuation), Type 2 (Style/Expression) and Type 3 (Idea Extension/Reduction) strategies involve evaluating and revising texts at a local level mostly within a sentence or a paragraph boundary. By contrast, Type 4 (Idea Addition/Deletion) and Type 5 (Reorganization) require evaluating texts at a comprehensive holistic level, reviewing a text in its entirety, taking its goal, audience, genre, and purpose into account. As such, the ‘improve’ group was expected to display a wider range of revision strategies encompassing more advanced strategies such as Type 4 and 5 in their revision repertoire compared to the ‘non-improve’ group. However, when the repertoire of revision strategies used by students throughout the semester was compared, no difference was found between the ‘improve’ and the ‘non-improve’ groups for both L1 and L2 language groups.

Specifically, in L1 group, three out of four students in the ‘improve’ group utilized all five revisions strategies and one (Doug) employed four types excluding the Type 5 (organizational change). In the ‘non-improve’ group, one student used all five and the other used four types, excluding Type 5. In L2 group, only one student (Megan) employed all five types of strategies and other three students utilized four types, ranging from Type 1 to Type 4 in the ‘improve’ group. Similarly in the ‘non-improve’ group, one student (Yves) used all five strategies and other two utilized four types from Type 1 to Type 4.

These results seem to suggest that the students in both the ‘improve’ and the ‘non-improve’ groups utilized similar repertoires of revision strategies but that L1 students have bigger repertoires of revision strategies than their L2 counterparts. However, all L2 students in the study employed meaning-level/global revision strategies. This makes a point of comparison with the revisions made by struggling writers (MacArthur et al., 1991). After examining revisions of struggling writers, the researchers suggested that these students may lack knowledge that revision entails global changes. Thus, although L2 writers may have a smaller repertoire of revision strategies than L1 writers, their revision repertoire seems larger than that of the struggling writers. The comparison of breadth of revision types employed by students between different groups is provided in Table 4.3.4.

Table 4.3.4 Comparison of Revision Repertoire

L1				L2			
Grp	Name	Revision Used	Strategies	Grp	Name	Revision Used	Strategies
High	Ethan	Type 1 ~ Type 5		High	Megan	Type 1 ~ Type 5	
High	Doug	Type 1 ~ Type 4		High	Harry	Type 1 ~ Type 4	
High	Peter	Type 1 ~ Type 5		High	Steve	Type 1 ~ Type 4	
High	Jordon	Type 1 ~ Type 5		High	Sally	Type 1 ~ Type 4	
Low	Wendy	Type 1 ~ Type 5		Low	Heather	Type 1 ~ Type 4	
Low	Amy	Type 1 ~ Type 4		Low	Yves	Type 1 ~ Type 5	
				Low	Jennie	Type 1 ~ Type 4	

4.3.1.2 Amount of Revision

L1 and L2 groups behaved differently on this variable. In L1, the ‘improve’ and the ‘non-improve’ groups made similar amounts of revisions across the semester while in L2 the ‘improve’ group made more revisions than the ‘non-improve’ group. The comparison of number of revisions by each group is provided in Table 4.3.5.

Table 4.3.5 Comparison of Number of Total Revisions by Group

L1				L2			
Improve Group		Non-Improve Group		Improve Group		Non-Improve Group	
	Revision		Revision		Revision		Revision
Ethan	38	Wendy	57	Megan	Re-write	Yves	72
Doug	56	Amy	42	Harry	92	Jenny	59
Peter	37			Steve	43	Heather	27
Jordon	49			Sally	71		
average	45	average	49.5	average	57	average	52.7

In L1, the ‘improve’ and the ‘non-improve’ groups made similar numbers of revisions over the semester (high = 45, low = 49.5). This is counter to the previous studies that suggest that generally more revision is associated with better writing and expert writers (Maynor, 1982; Monahan, 1982). However, the earlier studies only looked at a single session of writing and revision while the current study is a longitudinal study that includes three writing sessions. In fact, when the number of revisions made is broken down by session, some difference is found between the ‘improve’ and the ‘non-improve’ groups. In S1, the two groups made similar numbers of revisions (I = 18.33, NI = 18), but in S2 the ‘improve’ group made more revisions and the ‘non-improve’ group made considerably fewer revisions (I = 20.25, NI = 13.5). In S3, again, the two groups diverged with the ‘improve’ group’s revision falling drastically and the ‘non-improve’ group’s revision rebounding (H = 11.5, L = 18). These figures are presented in Table 4.3.6. These fluctuations in the amount of revisions may seem idiosyncratic at first but when it is examined together with the quality of draft 1 of papers, it reveals a point worth noting. The quality of the ‘improve’ group’s paper increased steadily across sessions whereas that of the non-improve group remained more or less the same across sessions. One possible explanation is that for the ‘improve’ group, the

consecutive increase in the amount of revisions in S1 and S2 seems to have had a cumulative effect on students' writing competency and contributed to the improvement of draft 1 to the point where revisions were not required as much in S3. On the other hand, the amount of 'non-improve' groups' revision seems not to have reached the threshold needed to bring growth in students' writing, and as a result, the papers of the 'non-improve' group continued to require meaningful revisions between draft 1 and draft 2.

In sum, these findings seem to suggest the existence of a threshold in the students' skills in utilizing peer feedback and making successful revisions. Those who attain this threshold through rigorous participation in repeated peer feedback sessions seem to gain development in their writing proficiency whereas the ones who fail to reach the threshold in the skills of utilizing feedback comments and making successful revisions appear to be unsuccessful in gaining growth in their writing competency. This idea is first raised by the present study and calls for further investigation in future studies using a larger number of participants.

Table 4.3.6 Changes in Amount of Revision across Sessions (L1)

Improve Group				Non-Improve Group			
	S1	S2	S3		S1	S2	S3
	Revisions	Revisions	Revisions		Revisions	Revisions	Revisions
Ethan	n/a	32	6	Wendy	18	24	15
Doug	14	21	23	Amy	18	3	21
Peter	14	16	7	average	18	13.5	18
Jordon	27	12	10				
average	18.33	20.25	11.5				

On the other hand, in L2, the 'improve' group generally made more revisions than the 'non-improve' group (Improve = 69, Non-Improve = 52.7). Even though the direction or the degree of growth in writing competency diverged as the semester progressed, the writing ability of students in two groups seem to have been similar to each other at the beginning of the semester based on the scores on their first essays. In terms of the number of revisions by session, little difference is found between the 'improve' and the 'non-improve' groups in S1

and S2. But this was partially due to the extensive revisions made by a student in the ‘non-improve’ group (Yves) from her incomplete first draft that was only half the length of another students’ paper. When excluding Yves, the average number of revisions of the ‘non-improve’ group is reduced to 19, which is considerably lower compared to the ‘improve’ group. In S3, the ‘improve group’ made considerably fewer revisions than S2, but the ‘non-improve’ group made almost no revisions. These data seem to suggest that similar to the L1 group’s case, the cumulative amount of revision and the reach of a threshold may account for the divide between the ‘improve’ and the ‘non-improve’ groups than the total number of revisions made.

Table 4.3.7 Changes in Amount of Revision across Sessions (L2)

<u>Improve Group</u>				<u>Non-I Group</u>			
	S1	S2	S3		S1	S2	S3
	Revisions	Revisions	Revisions		Revisions	Revisions	Revisions
Megan	21	47	24	Yves	32	40*	0
Harry	30	16	10	Jennie	34	23	2
Steve	15	27	29	Heather	11	15	1
Sally	43	12	3				
Average	27.25	25.5	16.5	average	25.67	26 (19)	1

* outlier with excessive revision from incomplete draft

4.3.1.3 Use of Type 4 Revision: Idea Addition/Deletion

In L1, the ‘improve’ group used Type 4 strategy less often than the ‘non-improve’ group in S1 (I = 5.67 vs. NI = 7.5) but more Type 4 revisions in S2 (I = 3.0 vs. NI = 1.5) and S3 (I = 4.75 vs. NI = 3.0). This breakdown by session is provided in Table 4.3.13. The detailed frequencies by revision type are provided in Table 4.3.8.

Table 4.3.8 Breakdown of Revisions by Type (L1)

<u>Improve Group</u>							<u>Non-ImproveGroup</u>						
	T 1	T2	T3	T4	T5	total		T 1	T2	T3	T4	T5	total
Ethan	1	19	12	3	3	38	Wendy	2	10	23	12	10	57
Doug	2	16	23	15	0	56	Amy	7	11	14	10	0	42
Peter	1	8	7	20	1	37							

Jordon	14	11	13	10	1	49									
Avg.	4.5	14	14	12	1.3	45	Avg.	4.5	11	19	11	5	49.5		

In L2 while Type 1 to Type 3 revision strategies did not show systematic difference neither across individuals nor between the groups, Type 4 (addition/deletion of new idea) strategy is employed much more frequently by the ‘improve’ group. With the exception of Yves, the ‘improve’ group utilized type 4 revision strategy almost four times as many as the ‘non-improve’ group (average of 5.75 vs. 1.5). Yves made eight Type 4 revisions in S2 alone due to the incomplete first draft of paper that had approximately half the length of other students’ papers. In addition, when the breakdown of use of Type 4 revisions by session is examined, it is shown that the two groups made similar amount of Type 4 revisions in S1 (I = 1.5 vs. NI = 1.6), the ‘improve’ group made more Type 4 revisions in S2 (I = 3.25 vs. NI = 0.5) and S3 (I = 1 vs. NI = 0), showing a trend similar to that found in L1 group. The frequency by revision type by group is provided in Table 4.3.9 and the breakdown by session is provided in Table 4.3.15.

Table 4.3.9 Breakdown of Revisions by Type (L2)

Improve Group							Non-improve Group						
	T 1	T2	T3	T4	T5	total		T 1	T2	T3	T4	T5	Total
Megan	11	34	39	6	2	92	Yves	12	24	23	10*	3	72
Harry	6	8	22	7	0	43	Jenny	8	24	25	2	0	59
Steve	26	28	11	6	0	71	Heather	6	12	8	1	0	27
Sally	16	15	22	4	0	57							
Average	14.8	21.3	23.5	5.8	0.5	65.8	average	8.7	20	18.7	4.33 (1.5)*	1	52.7

* an outlier with extensive revisions due to incomplete draft 1, () when excluding the outlier.

The difference in use of the Type 4 revision strategy by both improve groups may seem insignificant in face of a myriad of factors intricately affecting writing. However, it makes sense that adding an argument, a support, a warrant, a conclusion, or a combination of these to a paper can substantially improve the quality of the paper. While changes in

punctuation and grammar (Type 1), style (Type 2), and elaboration/reduction of ideas also affect the quality of paper, it is the strength/adequacy of the content that seems to have the strongest impact on the quality of the paper at least in case of the genre of academic writing. The data in the present study suggest that it was the weak content of the paper that was most responsible for low grades. It was far more detrimental to the quality of a paper than the problems of punctuation, grammar, and style. This observation is consistent with a prior study that investigated college undergraduates' revision (Butler et al., 2011) that states that global revisions are required to improve the quality of the essays.

An excerpt of example of students' revision with and without Type 4 strategy is provided in Table 4.3.10. The changes in Type 4 strategy are marked in ***arial black*** typeface in italic whereas the changes in other strategies are in **broadway** typeface. The comparison of different types of revisions types show that the magnitude of changes brought to the paper by Type 4 strategy is found to be clearly greater than that of other strategies.

Table 4.3.10 Revision with Type 4 Strategy

[Without Type 4 Strategy]

Draft 2

First, two authors Wendy Wright and Hugh McLachlan targeted different audiences in their articles. In "Cloning: Medical Miracle or Human Hubris?", Wright argues that human reproductive cloning creates a culture in which people are viewed as replaceable and controllable. She believes that the poor results obtained from animal cloning experiments should convince rational people that the "technology is unpredictable and dangerous." In her writing, she is especially intended to persuade the group of **women who are either currently or potentially interested in reproductive cloning not to try it. She addresses that reproductive cloning has a high rate of producing abnormal babies and it is dangerous not only to the baby but also to the prospective mother.** Hugh McLachlan, on the other hand, claims that fears about human cloning do not justify a ban on the practice. Instead, he maintains that cloning should be legalized to provide opportunities for couples who have no other way to conceive children. **While Wright specifically argues not to support reproductive cloning to the women who want to attempt it, McLachlan purposely speaks to the couples to try reproductive cloning for a chance.**

Even though two articles have different specific target audience, they have a common potential target audience. Both of the articles are probably attractive to many groups such as scholars and scientists who are involved or interested in human cloning research, people who either support or dismiss the idea of human cloning, and members of the Senate and House of Representatives who are eligible to vote on these issues. These groups of people believe that studying human cloning certainly has a lot of potential benefits **and** discoveries in science world. However, since human cloning is such a sensitive and unknown field, **they all think that** the U.S government should constitutionally draw certain guidelines.

[With Type 4 Strategy]

Draft 2

When writing an article the very first thing that the author must consider is the intended audience. It will be much easier to convince the intended audience if the author knows what they expect and how they think. A good author has to carefully analyse the audience to satisfy their expectations. The main audience who will be convinced by Cole's

article are people who are concerned about the environment and about future generations. This will generally be people who are a little older. As The Futurist magazine is mainly about how things are going to look in the future and how things have developed in recent times to make this future world, it is clear that it is the older generation that have interest in this subject. The main audience in Poore's article will be middle aged people who are unsure if garbage is a big problem and people who understand that there are environmental problems but are not ready to make drastic changes. As this article poses an opposing view to what people usually think, it will be a younger audience who will be interested in this. Since the audience for both these articles is different, the information provided to convince them should also be different.

Any article, for it to be convincing to the intended audience should provide enough logical information. The main line of reasoning in Coles's article is that there is no place where the garbage produced by the people in New York can be dumped and it is up to the people to work with the government to come up with a policy to reduce the dumping before it gets out of hand. **Coles provides a lot of evidence to support his point. According to him, "Fresh Kills Land fill has been the resting place for the garbage produced by New Yorkers for more than 60 years. It was once considered a bottomless pit but received its last load of waste on March 22nd, 2001, but New Yorkers haven't stopped producing wastes" (Coles). He makes this point to show people that the garbage produced by New Yorkers is monstrous and it is becoming hard for the government to find a good place to dump it. Coles also says, "The worldwide challenge will be to replace this throwaway economy with a future economy which reduces waste, reuses products and materials, and recycles" (Coles). He says this because he believes that the government should set up a policy to reduce this excess production and the people should support the government by following it.**

4.3.2. Feedback Incorporation

The amount and types of feedback received and incorporated by the 'improve' and the 'non-improve' groups were compared to examine the effect of the feedback received and incorporated on students' revision and writing improvement. More specifically, the amount of criticism comments and the amount of global-level criticism comments were tracked as they were found to be influential to the subsequent revision and the quality of the paper resulting from the revision from the preliminary analyses.

4.3.2.1 Feedback Received

The analyses of criticism and global criticism feedback received by essays from this study showed that generally, weaker papers invited more criticism and global criticism comments. As such, examining the criticism comments each group received offers some insight into the quality and characteristics of the group's writing. This information together with the data on the quantity of the criticism feedback received help reviewing incorporation of feedback with perspective.

4.3.2.1.1 Criticism and Global Criticism Comments Received

In L1, on average, the ‘improve’ group received more criticism comments and more global criticism comments than the ‘non-improve’ group (IC = 15.74 vs. NIC = 9; IGC = 9 vs. NIGC = 5). In L2, the ‘improve’ and the ‘non-improve’ groups received similar amount of criticism feedback (IC = 20.75 vs. NIC = 23.67), but the non-improve group received more global criticism comments (IGC = 5.25 vs. NIGC = 8). These figures are illustrated in Table 4.3.10 below.

These figures suggest that, the ‘improve’ group in L1 generally had weaker writing competency than the ‘non-improve’ group and the ‘improve’ and the ‘non-improve’ groups in L2 shared similar levels of writing skills but the ‘non-improve’ group may have been weaker on the global aspect of writing. This characterization is supported by the scores on their writing and their previous SAT and TOEFL scores. For the ‘improve’ group, the average scores of S1 writing, which is most reflective of the students’ general writing competency at the beginning of the semester, is much lower than that of the ‘non-improve’ group in L1 (I = 5.25 vs. NI = 9), and are similar to the two groups in L2 (I = 90 vs. NI = 91.3). The average SAT score for the improve group is moderately lower than that of the non-improve group (I = 577.5 vs. NI = 625) in L1 group, and the average TOEFL score for the improve group is slightly higher than that of the non-improve group (I = 98.8, NI = 93.5).

Table 4.3.11 Number of Criticism and Global Criticism Comments Received by Group

	L1		L2	
	Improve Grp	Non-Improve Grp	Improve Grp	Non-Improve Grp
Criticism Feedback Received	15.75	9.00	20.75	23.67
Global Criticism Received	7 (44.4%)	5 (33.3%)	5.25 (25.3%)	8 (33.8%)
S1 Writing Score	5.25	9.00	90.00	91.30
SAT/TOEFL Score	577	625	98.75	93.50

4.3.2.2. Feedback Incorporation

Analyses of the percentage of incorporation of criticism and global criticism feedback during revision show that in both L1 and L2 language groups, the improve group implemented higher percentages of both types of criticism feedback.

In L1 group, the students in the improve group incorporated 63.5% (10 comments) of criticism comments and 71.4% (5 comments) of global criticism comments during revision. In contrast, the students in the non-improve group incorporated a mere 22.2% (2 comments) of criticism comments and only 30% (1.5 comments) of global criticism comments.

Similarly, in L2 group, the students in the improve group incorporated 44.6% (9.25 comments) of criticism comments and 57.1% (3 comments) of global criticism comments during revision. On the other hand, the students in the non-improve group incorporated 32.4% (7.67 comments) of criticism comments and 41.6% (3.3 comments) of global criticism comments. The detailed breakdown of percentage of feedback incorporation by group is illustrated in Table 4.3.12.

Table 4.3.12 Percentage of Feedback Incorporation by Group

Feedback	L1		L2	
	Improve Group	NI Group	Improve Group	NI Group
Criticism Feedback Incorporated	10.00 (63.5%)	2.00 (22.2%)	9.25 (44.6%)	7.67 (32.4%)
Global Criticism Incorporated	5.00 (71.4%)	1.50 (30%)	3.00 (57.1%)	3.33 (41.6%)

These data suggest a few things. First, incorporating more criticism and global criticism comments seem to be associated with improvement in writing quality of students and possibly improvement in students' writing competency. Second, students' initial writing competency did not interfere with the level of feedback incorporation. In other words, the students who had weaker writing competency at the beginning of the semester (improve group in L1) incorporated more criticism and global criticism feedback during revision process than the students who may have had stronger writing skills (non-improve group in L1). Third, the amount of feedback one receives does not seem to affect the amount of

incorporation. In other words, receiving more criticism feedback does not necessarily mean that more will be incorporated. For instance, in L2, the non-improve group in L2 received more criticism and global criticism comments than the improve group but attained lower incorporation rate for both types of feedback.

What leads students to incorporate peer criticism feedback? Nelson and Schunn's (2009) empirically tested the effects of different types of feedback on incorporation of feedback and found that among various factors, understanding was the only significant mediator that determined implementation of feedback. In other words, students are more likely to incorporate feedback that they understand than the ones that they do not understand. Understanding was mediated by solutions, a summary of the performance, and the location of the problem indicating that understanding is more than simply detecting problems. It seems to be entailing what Flower et al. (1986) defined as *diagnosis*, that points the way to solution procedures. In short, lack of clear understanding, which entails not knowing what the problem is, where the problem is, or solutions that can fix the problem inhibit students' implementation of feedback they received.

The fact that the non-improve group in the study generated similar quantity and quality of feedback as the improve group suggest that their skills to detect problems in writing are similar to that of the improve group. As such, the cause for the low incorporation of peer feedback and the weak revisions resulting thereof seems to lie in students' inability to come up with solutions to fix the problem identified utilizing the right revision strategy. In other words, the students in the non-improve group in the present study seem to have the ability to identify problems of writing but are not apt to improve such problems, especially the global issues, during the revision process by adequately adding/deleting/moving texts; an ability that the students in the improve group seem to possess.

Butler & Britt (2011) posited that two potential source of difficulty in revision of argumentative essays are 1) poorly developed argument schema and, 2) a poorly developed global revision task schema, or schema on how to make global revision. The students in the non-improve group in the present study seem to have developed a fair level of global revision

task schema from the training from the class via a tutorial. This is supported by the fact that they were successful in utilizing global revision, albeit its low frequency. While data from the current study do not provide information on the level of argument schema developed by the students, developing skills to write argumentative essay was one of the course objectives, which is manifested by the final assignment given to the students: writing an argumentative essay. Since both groups of students were exposed to the same interventions and instructions on argument schema by the same instructor, there is little clue that directs to possible systematic difference between the two groups on this measure.

As Butler et al. state, building a strong argument schema and revision task schema is a minimum requirement for making global revisions. While attaining the schemas help students' performance in making global revisions, it does not solve all of the problems of student writing and revision (Butler et al., 2011). Even when students have attained the schemas, as in the present study, some students still find it difficult to make effective revisions.

One such student is Jennie in the current study. Jennie was one of the most enthusiastic students in the class who participated in all three peer review sessions, reviewed all twelve papers during the semester, generating far more feedback comments compared to the class average. Her feedback also had the features of helpful feedback: high percentage of criticism comments focusing on global features, specific and detailed comments with a high percentage of solutions. Looking at the feedback generated alone, it is hard to see her categorized in the non-improve group. Nevertheless, she was not very successful in making substantial global changes during the revision process and failed to achieve improvement between her drafts as well as across the assignments over the semester. While there may be several cause for Jennie's weak revision, a review of her papers across sessions show high re-use of sentences across the papers, revealing text encoding, i.e. sentence construction as one of Jennie's problems. For L2 writers, competency in text generation does not come naturally with age and previous studies on L2 writing have identified text generation as one of the key challenges for writing in target language (de Larios et al., 2006; Zimmermann, 2000).

McCutchen (2000) contemplates that writing expertise depends on the development of 1) fluent language generation processes and 2) extensive knowledge relevant to writing (e.g., topic knowledge, genre knowledge). She applies the concept of long-term working memory (LT-WM) developed by Ericsson and Kintsch (1995), and explains writing as an orchestrated process of connecting and utilizing short-term working memory (ST-WM); which concerns fluent language generation process, and LT-WM; which deals with extensive knowledge relevant to writing. Using this approach, McCutchen claims that novice writers, with lack of fluency in language encoding processes, remain limited by ST-WM capacity and are not able to retrieve information from LT-WM. She contends that when language encoding processes become sufficiently fluent, writers can transcend the limits of ST-WM and capitalize on LT-WM. In such cases, without deep topic and genre knowledge developed in LT-WM, writers can lose access to LT-WM resources and have to resort to less mature strategies. In short, to become skilled writers, one needs to develop fluent text generation process as well as extensive knowledge on writing, such as topic and genre knowledge to move beyond the limits of short-term working memory and capitalize on the resources of long-term working memory.

Applying McCutchen's model, Jenny seems to have developed sound topic and genre knowledge in LT-WM but her insufficient fluency in text generation confined her to the boundary of ST-WM and impeded her from writing a paper that capitalized on the topic and genre knowledge. Review of her drafts show that from draft 1 of S1 to draft 2 of S2, she has frequently employed sentence re-use strategy, which is often the result of borrowing sentences from the source text and her previous drafts (all four drafts were on the same topic). However, in draft 1 of S3, a marked difference is observed in the style, format and fluency of the text compared to her earlier drafts. In addition, the content of the paper does not satisfy the requirement of an assignment in which an argumentative essay for specific readers was asked for. Taken together, it seems highly possible that she copied a portion of an article and submitted it for peer review. As one can expect, the feedback from her peers was mostly negative and she re-wrote the paper from scratch in draft 2, which became a very

short, unelaborated, weak essay, which suggests her weak text generation process. McCutchen calls for empirical studies to fill in the details concerning the specific theoretical predictions she made. Her study mainly targets L1 writers who have developed fluency in text generation yet have not fully developed the writing knowledge. It would be interesting to test her theory by investigating whether the writers who struggle in making global revisions, despite having developed strong writing knowledge, can improve their performance by developing fluent text generation processes.

4.3.3 Feedback Provided

The quantity and the quality of feedback provided by the improve and the non-improve groups were compared using the coding scheme introduced in Chapter 3. Detailed-level categories were collapsed to create categories that could represent the quality of feedback based on the helpfulness to the writer reported in a previous study (Cho et al, 2006). The collapsed categories of feedback compared between the two groups are given below.

- Total number feedback produced measured by idea units
- Percentage of total feedback broken down into praise, criticism, and reflection
- Percentage of criticism feedback addressing macro vs. micro issues (out of total feedback)
- Percentage of all feedback categorized into general vs. specific comments
- Criticism feedback sub-categorized into problem, solution and problem+solution

Overall, for both the L1 and the L2 groups, the feedback comments generated by the ‘improve’ and the ‘non-improve’ groups shared similar profile. The feedback of the non-improve group contained slightly more criticism comments, addressed more global features of writing and was more specific, which are the features of more helpful feedback. In terms of quantity, the high and the low groups in L2 produced similar amount of idea units but the improve group generated more idea units in L1 group. The detailed breakdown of feedback in each category is provided in Table 4.3.13 and Table 4.3.14 for each language group.

Table 4.3.13 Comparison of Feedback by Improve and Non-improve Groups (L1)

Grp	Name	idea unit	prse	crtc	rflctn	mcro	micro	gnral	spcfc	prbm	soltn	p+s
High	Ethan	122	50.82	36.89	9.84	26.23	10.66	39.34	48.36	9.02	12.30	15.57
High	Doug	132	32.58	55.30	8.33	33.33	21.97	53.03	34.85	23.48	16.70	15.15
High	Peter	71	56.34	26.76	16.90	19.72	7.04	66.20	16.90	16.90	5.63	4.23
High	Jordon	43	37.21	39.53	16.28	25.58	13.95	55.81	20.93	18.60	2.33	18.68
	Avg.	92	44.23	39.62	12.84	26.22	13.40	53.60	30.26	17.00	9.24	13.41
Low	Wendy	65	33.85	46.15	20.00	27.69	18.46	38.46	41.54	29.23	9.23	7.69
Low	Amy	79	25.32	67.09	6.33	39.24	27.85	51.90	40.51	22.78	26.60	17.72
	Avg.	72	29.58	56.62	13.17	33.47	23.16	45.18	41.03	26.00	17.92	12.71

Table 4.3.14 Comparison of Feedback by Improve and Non-improve Groups (L2)

Grp	Name	idea unit	prse	crtc	rflctn	Mcro	micro	genrl	spcfc	prbl m	soltn	p+s
High	Megan	147	46.26	46.26	9.52	34.01	12.25	61.90	30.60	29.25	7.48	9.52
High	Harry	120	41.67	43.44	9.17	30.83	12.50	66.33	21.67	30.00	3.33	10.00
High	Steve	111	45.95	38.74	5.41	26.13	12.61	77.48	7.21	29.73	6.31	2.70
High	Sally	79	44.3	31.65	16.46	26.58	5.06	55.70	20.25	15.19	6.33	10.13
	Avg.	114.30	44.55	40.02	10.14	29.39	10.61	65.35	19.93	26.04	5.86	8.02
Low	Heather	100	46.00	37.00	15.00	30.00	7.00	51.00	32.00	19.00	9.00	9.00
Low	Yves	120	44.17	42.50	16.67	29.17	13.33	60.00	26.67	22.50	5.83	14.17
Low	Jennie	128	39.06	52.34	7.81	46.09	6.25	40.63	50.78	17.19	19.50	15.63
	Avg.	116	43.08	43.95	13.16	35.09	8.86	50.54	36.38	19.56	11.44	12.93

A point worth noting is when looking at the proportion of feedback, the non-improve group seems to have produced more helpful comments with higher proportion of criticism (56.62% vs. 39.62%), specific (41.03% vs. 30.26%), and solution (17.92% vs. 9.24%) comments. But when the comments are compared in terms of absolute number of idea units, the improve group produced more criticism comments (40.7 vs. 21.3 idea units), similar amount of specific (27.8 vs. 29.5) and solution (8.5 vs. 12.9) comments. This finding may be suggesting that the absolute number of feedback comments by type may be a better indicator of the quality of feedback generated than the proportion of feedback by type. But considering the small sample size used in the present case study, it needs to be interpreted with caution.

In a peer feedback study, Lundstrom & Baker (2009) report that when students engage in only one-side of peer feedback activity, either giving or receiving feedback, giving feedback is found to be more effective in improving writing. Based on this finding, it was expected that the improve group would produce feedback higher in quantity and quality than the non-improve group. Surprisingly, the data from the present study demonstrate that, for both L1 and L2 groups, when students are engaged in both giving and receiving peer feedback, they vary little in producing the feedback comments. Under feedback sessions where the format of feedback is relatively controlled and the cohort of students participating in feedback share similar educational background, such as freshmen enrolled in the same tertiary institution, the amount and the types of feedback generated seem to be more bound by the characteristics of the text than that of the reviewer. The overall similarity in the profile of feedback generated between the improve and the non-improve groups imply that the students in the improve and the non-improve groups in the study share parallel skills in evaluating and detecting issues in writing and verbalizing it into feedback comments. From this, it can also be inferred that the students in this study received fairly similar quality of feedback comments. Nevertheless, they diverged in how they progressed in developing writing skills over a semester. Based on this data, it can be inferred that the ability to accurately evaluate other's writing may be a necessary condition but not a sufficient condition for good writers. On the flip side, receiving high quality feedback may be a necessary condition to improve

writing quality but it not a sufficient condition to improve writing quality. At least for the students in the present study, the way in which they engaged in providing peer feedback; frequency, amount, types of feedback; in itself did not directly affect improvement in their writing competency.

4.3.4 Longitudinal Effect of Feedback and Revision on Writing

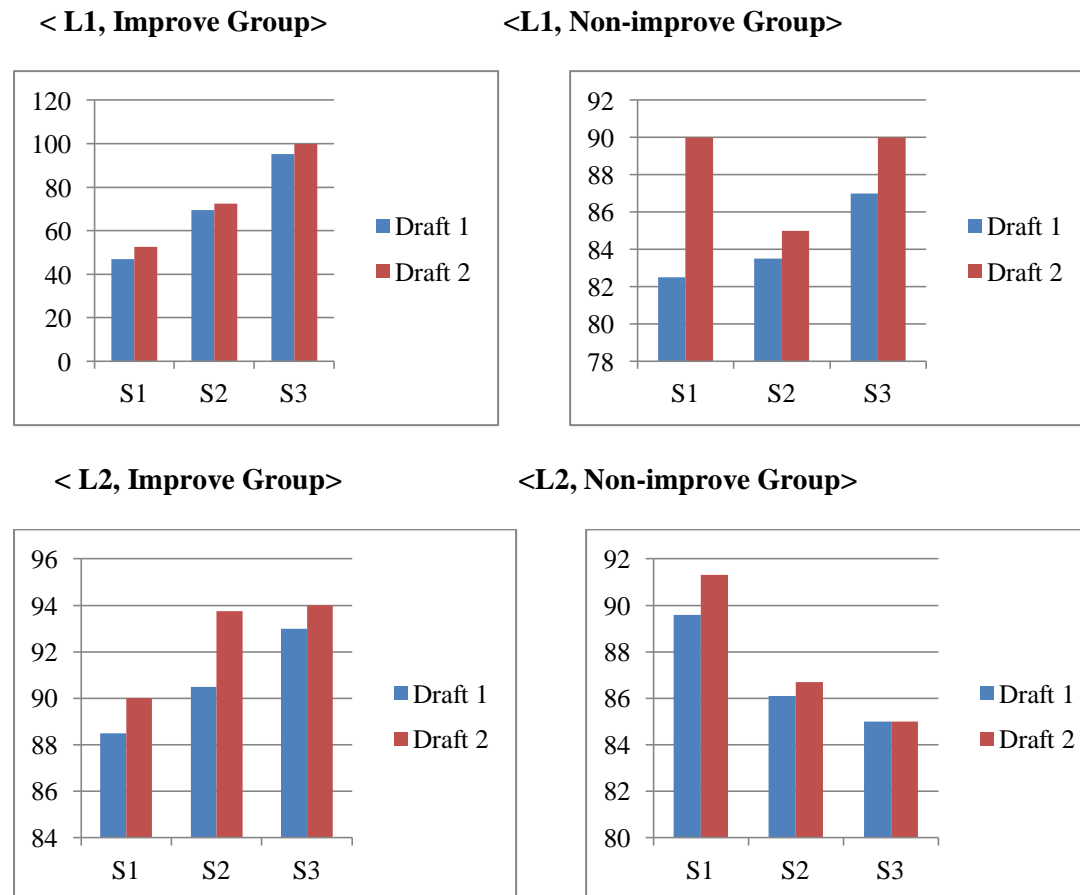
In this study, the students in the improve group were the ones who gained consecutive increase in their writing scores over the semester and the students in the non-improve group were the ones whose scores on their writing either stayed the same or decreased across the writing sessions. In order to illustrate the degree and the direction of changes in students' writing competency over six drafts that the students submitted over the semester, the quality of six drafts were compared. This comparison provides a snapshot of the interim states of students' writing proficiency across the six drafts of essays the students' wrote over a semester. As a representation of essays' quality, the scores awarded by the instructors were used for the second drafts. For draft 1, due to absence of instructor grades, virtual scores were devised based on the scores on draft 2 and the magnitude of revisions made. Since the use of Type 4 revision is best predictor of improvement in writing quality, the scores of draft 1 was calculated by deducting the number of Type 4 revisions made between draft 1 and draft 2. The virtual and actual scores of six drafts of essays resulting from this calculation are provided in Table 4.3.15, and illustrated in Figure 4.3. These scores are only meant to represent the relative quality of students' essays produced over a semester and do not intend to accurately depict the quality of each paper nor the level of students' writing proficiency in absolute terms.

Table 4.3.15 Virtual Scores on Six Drafts of Essays

Essays	L1		L2	
	Improve Group	Non-Improve Group	Improve Group	Non-Improve Group
Draft 1, S1	46.8	82.5	88.5	90.0
Draft 2, S1	52.5	90.0	90.5	93.8

Draft 1, S2	69.5	83.5	93.0	94.0
Draft 2, S2	72.5	85.0	89.6	91.3
Draft 1, S3	95.3	87.0	86.1	86.7
Draft 2, S3	100.0	90.0	85.0	85.0

Figure 4.3 Changes in Virtual Scores on Six Drafts of Essays over Time



The scores show more or less similar findings for both language groups, especially for the improve groups. The improve groups experienced a steady improvement in writing between the drafts as well as across the sessions. This steady increase in writing quality appears to indicate accumulation of growth. In other words, the growth gained from draft 1 to draft 2 of S1 is reinforced and cumulated to the growth from S2, which then is cumulated to the gains from S3. It seems that the students who witnessed growth in writing proficiency over a semester achieved growth between drafts within a paper through stronger revision

mediated by higher incorporation of peer feedback. Then, the effect of that growth seems to have extended beyond a single paper, building on to their general writing ability thereby affecting the writing of the following assignments.

Another point worth noting is that the improve group made more revisions in S2. The analysis of longitudinal changes in the peer feedback of the larger group (N = 38) discussed in Chapter 4.2 shows that the students who participated in the current study produced less idea units, less words, and less criticism comments across sessions. It seems contradictory that the students incorporated higher percentage of peer feedback when the comments actually became less specific and less helpful. One explanation can be that the students' perception and trust towards peer feedback was positively influenced by the peer feedback experience in S1 to the degree that encouraged them to accept more peer feedback despite its lower quality in S2. In addition, repeated practice of reviewing and evaluating peers' paper may have enhanced their ability to detect issues in their own paper and encouraged them to accept more feedback from peers.

Conversely, students in the non-improve group achieved high scores in S1 but their scores declined over the semester as they incorporated low rates of peer feedback and made insignificant changes during the revision process. For L1 group, the quality of draft 1 shows slight increase across sessions, although small, but for L2 group, the quality of draft1 steadily went down over the semester. The numbers of total and Type 4 revisions as well as the rate of peer feedback incorporated of the L2 non-improve group were all lower than that of the improve group, excluding Yves who was an outlier. It seems that these students' failure in achieving considerable improvement between the drafts was extended across assignments resulting in little growth over the semester.

Although based on a small case study, the finding from the present study provides a small clue that 1) the percentage of criticism feedback incorporated, the amount of total and Type 4 revisions made positively affect the development of students' writing proficiency.

2) the repeated experience of peer feedback and revision has a longitudinal effect of improving students' performance in incorporation of peer feedback and making more

significant revisions, and 3) the gains in writing attained from making revisions can cumulatively add to the general writing competency of students, enhancing their writing of following assignments. The detailed number of total revisions, Type 4 revisions, and writing scores of each group by student is provide in Table 4.3.16 and Table 4.3.17.

Table 4.3.16 Scores and Revisions by Type across Sessions (L1)

Improve Group									
	S1			S2			S3		
	Total Revision	Type 4	Score	Total	Type 4	Score	Total	Type 4	Score
Ethan	n/a	n/a	2	32	3	8	6	0	11
Doug	14	4	10	21	2	10	23	9	12
Peter	14	7	4	16	7	5	7	6	10
Jordon	27	6	5	12	0	6	10	4	9
average	18.33	5.67	5.25	20.25	3.00	7.25	11.50	4.75	10.50

Non-improve Group									
	S1			S2			S3		
	Total Revision	Type 4	Score	Total	Type 4	Score	Total	Type 4	Score
Wendy	18	3	8	24	3	7	15	6	8
Amy	18	12	10	3	0	10	21	0	10
average	18	7.5	9	13.5	1.5	8.5	18	3	9

Table 4.3.17 Scores and Revisions by Type across Sessions (L2)

Improve Group									
	S1			S2			S3		
	Total Revision	Type 4	Score	Total	Type 4	Score	Total	Type 4	Score
Megan	21	3	92	47	3	95	24	0	94
Harry	30	2	85	16	1	94	10	4	94
Steve	15	1	92	27	5	94	29	0	95
Sally	43	0	91	12	4	92	3	0	93
average	27.25	1.5	90	25.5	3.25	93.75	16.5	1	94

Non-improve Group									
	S1			S2			S3		
	Total revision	Type 4	Score	Total	Type 4	Score	Total	Type 4	Score

Yves	32	3	86	40	8*	78	0	0	75
Jennie	34	2	94	23	0	93	2	0	90
Heather	11	0	94	15	1	89	1	0	90
average	25.67	1.67	91.3	26	0.5	86.67	1	0	85

** excluded from average as an outlier*

CHAPTER 5. CONCLUSION

Peer review and feedback have long been used in the L2 classrooms but with reservations and limited roles. Behind this lie perceptions shared by teachers and learners that L2 learners' limited language capacity bounds the ways in which peer feedback can be used in the L2 setting. Unlike the active role peer feedback plays in L1 classrooms, the use of peer feedback in L2 writing classrooms has been limited to a peripheral role. Nevertheless, studies have not yet closely compared the feedback comments of L1 and L2 students and the effects feedback has on each group's revision. The present study investigated the characteristics and the usefulness of peer feedback comments generated by L2 learners in comparison with that of L1 learners and the longitudinal aspects of peer review and feedback. In particular, it closely examined how the students' peer feedback comments change over time through repeated practice, and how the peer review activity affects students writing improvement in a long term. In this concluding chapter, I first summarize the findings from the series of analyses as they relate to the research questions. Second, the theoretical, methodological, and pedagogical implications associated with these findings are presented. Third, the limitations embedded in the study that weaken the generalizability of the results are discussed. Finally, directions for the future research investigating the topic of peer feedback and writing are proposed.

5.1 SUMMARY OF THE FINDINGS

5.1.1 Research Question One

The first research question is how the peer feedback comments generated by the L2 learners differ from or coincide with the peer feedback comments made by L1 learners when performed under comparable environment using SWORD, an online peer review tool.

First, the overall participation level of the students in each group and the productivity of feedback generated were compared. Contrary to common belief that L2 reviewers are less productive than L1 readers, data showed that L2 learners on average produced more idea

units ($L2 = 93.2$ vs. $L1 = 70.9$) and words ($L2 = 1,049$ vs. $L1 = 950$) over the semester than their L1 counterparts. In addition, L2 learners had higher participation level in the peer review activity than L1 students based on the number of essays reviewed and the number of peer review sessions participated. When the amount of idea units and words generated per essay were compared between the two language groups, L2 learners were found to produce more idea units and words than L1 learners, although not significantly more. L1 learners used significantly more number of words per idea unit ($L1 = 13.35$ vs. $L2 = 11.29$) than L2 students when giving feedback, suggesting that L1 reviewers are more verbose than L2 reviewers.

Second, the feedback comments of the two groups were compared using feedback categories. When the feedback comments were compared at high level, the analysis showed no statistical significant difference between the two groups in terms of distribution across the praise, criticism, and reflection comments. In line with the previous SWORD study (Patchen et al., 2009), both groups produced praise comments the most ($L1 = 46.19\%$, $L2 = 46.41\%$), followed by criticism comments ($L1 = 38.73\%$, $L2 = 38.36\%$), and reflection comments ($L1 = 14.03\%$, $L2 = 11.6\%$). No significant difference between the two groups was found in terms of feedback specificity (general vs. specific) and scope (macro vs. micro). On the other hand, a detailed level analysis of the criticism category showed that the L2 learners gave significantly more *problem-only* comments, a type of feedback that explicitly depicts the problem of a paper without providing solution, than the L1 learners. Conversely, L1 students made more solution-only commentaries than their L2 counterparts, close to reaching a statistically significant level of difference. According to Patchen and her colleagues, this difference may be attributed to the difference in the two groups' orientation: L1 students seem to be more solution-oriented and L2 students more problem-oriented when they review their peers' essays. This difference in orientation may stem from the varying cultural and social background of the students in the two groups, as well as the different levels of experience they had on peer review and feedback prior to the current study.

In sum, the comparison of the two groups' feedback comments on 16 different features represented in the feedback categories show that the feedback comments generated by the two groups of students are similar in all features measured except for the explicit depiction of problems and solutions. In terms of quantity, the only difference between the two groups was that L1 students used more words to discuss a unit of idea. All in all, these results seem to suggest that L1 and L2 students generate parallel peer feedback comments in quantity and quality when they engage in peer review under similar environment.

In addition to the categorical differences that reflect the function and the degree of helpfulness of the feedback comments, textual features of the comments were compared. The analysis of textual features showed that two groups of learners diverged on three aspects. First, the two groups varied on their use of modal verbs. Particularly, when providing solutions, L2 reviewers tended to employ obligation/necessity modals, such as 'should' and 'have to,' while L1 learners used possibility/ability modals, such as 'can,' 'may,' or 'could.' Second, the two groups differed in the use of the modal 'can.' While the modal was used mostly as a politeness marker by L1 students to not be offensive or authoritative to their peers, L2 learners often used 'can' to denote possibility and to indicate possible room for improvement. Third, when providing criticism, L1 learners often used mitigation language, such as giving meaningless praise or downplaying their criticism as a face-saving strategy but L2 students rarely used these.

These differences can be explained by the face-saving model (Brown and Levinson, 1978, 1987) and face-threatening acts (FTAs). Acts that threaten face include advice, orders, requests, suggestions and warnings and students' solution comments, being suggestions, are one of FTAs. Studies have shown that different politeness strategies are used to mitigate the effects of FTAs and that the politeness strategies and individual speech acts may vary from one language/culture to another and between a native versus a non-native speaker. While these differences may not directly affect the function and/or the helpfulness of the feedback comments, they may create confusion, cause offensiveness, or interfere with the effectiveness of the peer feedback in classrooms comprised of mixed population of L1 and L2, which is

common in today's multi-cultural educational environment. Accordingly, these textual differences between the two groups need to be acknowledged by the instructors employing the peer review method in those classrooms so that any possible misunderstanding or interference during the peer review process is minimized.

5.1.2 Research Question Two

The second research question asked how the peer feedback comments of L1 and L2 students change or develop over a semester. To answer this question, the feedback comments were coded for each session and compared across sessions on the distribution of feedback across the feedback categories. Then, the changes were compared between the groups to identify how the two groups' feedback comments converged or diverged over time.

The analyses were conducted in layers. First, repeated MANOVA was employed to check the changes at high level categories of praise, criticism, and reflection. This revealed that both language groups gave more praise comments and less criticism and reflection comments over time as the semester progressed, although the changes were not at a statistically significant level. In other words, at this high level, the percentage of feedback remained more or less the same across the three peer review sessions throughout the semester for both groups. This finding is contrary to the expectation that repeated practice will lead to improvement in the types of feedback generated: meaning increase in criticism and reflection and decrease in praise. One possible explanation for these counter-intuitive results is that the students were not recognized/rewarded to produce more helpful commentaries. Thus, unlike the expectation that increased time on task will lead to improvement, the students came to provide feedback that were quicker and easier to give and shifted their efforts, attention, and time to the tasks that were more recognized, rewarded, or enforced. Accordingly, increased time on task will not lead to better performance unless proper reward or control for the task is also implemented. This finding offers an important pedagogical implication with regards to designing multiple peer review/feedback sessions in classrooms.

When the changes in the sub-categories of criticism are compared, both the effect of time and the interaction between time and group were insignificant. These results suggest that

the two groups did not change significantly in the percentages in the criticism sub-categories of problem-only, solution-only, and problem+solution, nor did they change in different directions for what changes there were. Similarly, analyses on the specificity and scope of feedback showed that the feedback comments of the two groups did not change significantly on these features over the semester.

Although not statistically significant, there were some changes that took place in the feedback commentaries of the students in both groups. The general tendency was that the feedback comments became less rigorous over time. The feedback comments of both groups became less specific and addressed fewer global issues. These changes again seem to indicate that the students exerted less effort in generating helpful feedback. One notable—although statistically insignificant—change is that over time, L1 students came to generate more problem-only comments and less solution-only comments. In contrast, L2 students produced more solution-only comments and less problem-only commentaries. These findings suggest that although engaging students to repeated practice of peer review and feedback did not bring significant changes to the feedback profiles of the students, it may have helped students shift their orientation to the areas they have been less attentive to previously and as a result help students develop more balanced orientation for problems and solutions.

The student survey was analyzed to capture the changes in students' perception and behavior regarding peer review and feedback. Both L1 and L2 students reported that they became more interested in providing and receiving peer feedback over time, and that both the feedback they provided and received became more helpful in their revision over time. When asked about the changes in the incorporation of peer feedback over time, the students in both groups answered that they accepted more feedback over time. L2 students reported higher percentage of peer feedback incorporation than L1 learners, contrary to earlier studies reporting L2 learners' distrust for peer commentary (Carson and Nelson, 1996; Zhang, 1995). In another set of questions inquiring about the feedback skills, students in both groups reported growth in understanding peer comments better, identifying problems in paper, devising solutions for problems, and providing more accurate feedback. The survey responses

suggest that the repeated exposure to peer feedback and revision helped students understand and utilize the feedback from their peers better, with increased engagement in peer review activity, even though the feedback comments generated did not improve.

5.1.3 Research Question Three

This research question explores the longitudinal relationship between the ways students engage in peer review and revision, and their writing development reflected in the writing scores. To pursue this question, two subgroups of sample student cases were selected from each language group for comparison: the improve group, whose writing scores improved over the semester, and the non-improve group, whose writing scores stayed the same or declined across the sessions. These subgroups' performance on writing, peer feedback, and revision activities were analyzed and compared around five themes to identify links between the peer review/feedback activity and writing development: types of revisions made, amount of revisions made, peer feedback received, peer feedback provided, and peer feedback incorporated.

To examine the types of revisions made, changes made to the text between the drafts were tracked and categorized into five revision types/strategies ranging from Type 1 to Type 5: changes in surface features such as grammar (Type 1), changes in style and expression (Type 2), extension/reduction of part of idea (Type 3), addition/deletion of whole idea (Type 4), and changes in organization (Type 5). First, the breadth of revision strategies employed over the semester between the 'improve' and the 'non-improve' groups were compared. Analyses showed that in both language groups, the improve and the non-improve groups shared similar breadth of revision strategies, with L1 students employing slightly bigger revision repertoire than L2 learners.

Second, the amount of total revisions made over the semester, as well as the breakdown by session was compared between the improve and the non-improve groups. In L1 group, the amount of total revisions did not vary between the improve and the non-improve groups ($I = 49.5$ vs. $NI = 52.7$) but the breakdown by session showed a clear

difference in the trend between the improve and the non-improve groups from S1 to S2 to S3. The improve group made more revisions in S2 than S1 and the least in S3 ($S1 = 18.33$, $S2 = 20.25$, $S3 = 11.5$), but the non-improve group made less revisions from S1 to S2 but the number of revisions rebounded in S3 ($S1 = 18$, $S2 = 13.5$, $S3 = 18$). Review of the quality of draft 1s across sessions revealed that the improve group wrote better quality draft 1s over time, implying that their writing required less revision between drafts from S1 to S2 to S3. This explains the low amount of revision made in S3. However, the group made more revisions in S2 than in S1, even though the quality of draft 1 in S2 was better than the draft 1 of S1. Moreover, this took place while the quality of feedback in S2 actually deteriorated from S1 as revealed from the analysis of feedback profile (Section 4.2). These findings taken together with the students' survey responses suggest that the students in the improve group became better at understanding and utilizing peer feedback comments even in the face of their quality decline and made more revisions in S2 than S1, indicating learning effect from engaging in peer feedback and revision multiple times across sessions. In addition, this consecutive increase in the amount of revisions is speculated to have had a cumulative effect on students' writing ability over time, leading them to achieve higher quality draft 1s across sessions.

By contrast, the amount of revisions made by the non-improve group decreased from S1 to S2, presumably falling short of reaching a threshold required to bring growth in students' writing. As a result, the quality of draft 1s of the non-improve group did not show improvement across sessions, continuing to leave much room for improvement and revising in S3.

In L2 group, the improve group generally made more revisions than the non-improve group ($I = 69$ vs. $NI = 52.7$). However, when the number of revisions were broken down by session, there was little variance found between the improve and the non-improve groups. But when the outlier (Yves) who made extensive revisions in S2 due to incomplete draft 1 was removed, the performance of the two groups showed a contrast. The improve group made similar amount of revisions in S1 and in S2, but less in S3, similar to the trend shown in

the improve group in L1 ($S1 = 27.25$, $S2 = 25.5$, $S3 = 16.5$). Conversely, the non-improve L2 group showed a steady decline in the amount of revisions made from S1 to S2 to S3 ($S1 = 25.67$, $S2 = 19$, $S3 = 1$).

In sum, these findings seem to suggest the existence of a threshold point in the students' skills in utilizing peer feedback and making successful revisions. Those who attain this threshold through rigorous participation in repeated peer feedback sessions seem to gain development in their writing proficiency whereas the ones who fail to reach the threshold in the skills of utilizing feedback comments and making successful revisions appear to be unsuccessful in gaining growth in their writing competency.

Third, the use of Type 4 revision emerges as a factor that differentiated the improve and the non-improve groups. In L1, although the number of Type 4 revisions used per student over the semester is similar between the improve and the non-improve groups ($I = 12$ vs. $NI = 11$), the breakdown by session shows that the non-improve group's use of Type 4 revision dropped significantly from S1 ($I = 5.67$ vs. $NI = 7.5$) to S2 ($I = 3$ vs. $NI = 1.5$) to S3 ($I = 4.75$ vs. $NI = 3$). In L2 group, with the exception of Yves, who made extensive revisions in S2 due to incomplete draft 1, the improve group made far more Type 4 revisions over the semester than the non-improve group. In addition, the breakdown by session also shows that while the two groups made similar amount of Type 4 revision in S1 ($I = 1.5$ vs. $NI = 1.6$), the improve group made much more Type 4 revisions than the non-improve group in S2 ($I = 3.25$ vs. $NI = 0.5$), and in S3 ($I = 1$ vs. $NI = 0$). These findings suggest that both in L1 and L2 context, the student writers who engage in more global revisions (Type 4) achieve more growth in writing proficiency over a semester.

Fourth, the relationship between the feedback received, the feedback incorporated, and improvement in writing was investigated by comparing the improve and the non-improve groups. The analyses showed that among other factors, the percentage of the critical feedback and the global-level critical feedback comments incorporated during revision divided the improve and the non-improve groups both in L1 and L2 language groups. In both language groups, the improve subgroups incorporated higher percentages of criticism comments ($L1I =$

63.5% vs. L1NI = 22.2%; L2I = 44.6% vs. L2NI = 32.4%), and higher portions of global-level criticism feedback (L1I = 71.4% vs. L1NI = 30%; L2I = 57.1% vs. L2NI = 41.6%) than the non-improve groups. When the amount of criticism and global-level criticism feedback received by each group is examined in relation to the incorporation level, no direct relationship was found. The improve group in L1 received more criticism and global-level criticism comments than the non-improve group (IC = 15.75, IGC = 7 vs. NIC = 9, NIGC = 5). On the other hand, in L2, the improve group received less criticism and global-criticism comments than the non-improve group (IC = 20.75, IGC = 5.25 vs. NIC = 23.67, NIGC = 8). Review of the students' essays suggests that the amount of criticism and global-criticism comments received are more related to the quality of the paper being reviewed, with weaker papers inviting more criticism and global-criticism comments. Thus, the percentage of incorporation of criticism feedback seems to be a better predictor of the improvement made to the draft than the absolute number of criticism comments incorporated. Also, the percentage of feedback incorporation was not interfered by the level of students' writing proficiency. In the L1 group's case, students who displayed weaker writing ability at the beginning of the semester incorporated higher percentage of criticism feedback than the students who displayed strong writing proficiency at the beginning of the semester. The large magnitude of revisions made, mostly influenced by the incorporation of criticism feedback, appears to have led the weaker writers to gain growth in their writing ability to the level that exceeded the stronger students over the semester.

Fifth, comparison of peer feedback provided by the improve and the non-improve groups revealed little variance between the two groups in both language groups. This finding suggests that when a cohort of students produces peer feedback using relatively controlled format, the quantity and quality of feedback produced are more bound by the characteristics of the text than that of the reviewers. The similar quantity and quality of feedback generated by the improve and the non-improve subgroups hint that the ability to evaluate others' writing may be a necessary condition but not a sufficient condition to become a good writer.

Lastly, the quality of six drafts—draft 1 of S1, draft 2 of S1, draft 1 of S2, draft 2 of S2, draft 1 of S3, draft 2 of S3--written throughout the semester by the improve and the non-improve groups was compared to reveal a snapshot of the interim states of students' writing ability. This aimed at investigating the developmental paths of the improve and the non-improve groups. For six drafts of paper, scores awarded by instructors were used as measures representing the quality of three draft 2s of S1 to S3. Due to the absence of instructors' scores awarded to draft 1s, the virtual scores were calculated by taking the scores awarded by the instructors to draft 2s minus the number of Type 4 revisions made between draft 1 and draft 2. The analyses show that the improve groups of both language groups attained steady growth in writing scores: the quality of their draft 1s increased from S1 to S2 to S3. But the draft 1s of the non-improve subgroup gained only minimal growth across sessions in L1 group, and even declined over the semester in L2 group. The non-improve groups had lower number of total revisions, lower number of Type 4 revisions and lower rate of incorporation of criticism and global criticism comments throughout the semester compared to the improve group. These findings indicate that the amount of Type 4 and total revisions made, along with the incorporation of criticism comments, may reflect the development of students' writing competency.

Put together, the findings from the analyses in this section suggest the following:

- a. Repeated experience of peer feedback and revision has a long-term effect of improving students' performance in incorporating peer feedback comments and making more successful revisions.
- b. Incorporation of criticism and global criticism comments, number of total revisions made, and the number of Type 4 revisions made positively affect the development of students' writing proficiency.

- c. The gains in writing attained from making revisions between the drafts adds to the general writing ability of students and extends to writings beyond the current paper.

5.2 IMPLICATIONS

5.2.1 Theoretical Implications

The present study has three theoretical implications. First and foremost, this study bridges the gap existing between the peer feedback and composition studies in the L1 and the L2 fields. Few scholars have attempted to compare the peer review and feedback activities of the two learner groups. Rather, the myriads of findings discovered about the role, benefits, and efficacy of the method of peer review and feedback are confined to their respective field and rarely benefit the scholars studying the other field. This study, by using SWORD to compare the peer feedback comments generated by L1 and L2 writers under comparable settings and conditions, attempted to minimize the variables embedded in the existing studies to enhance comparability of the study results. The study found that the two groups of students generally behave similarly in generating peer feedback comments. This finding lays a foundation on which the findings on the topic of peer feedback and writing segregated in the respective fields of L1 and L2 contexts can be related, connected, compared, and contrasted. In addition, the present study employed SWORD, an online peer review tool adopted by a series of peer review studies conducted in recent years in the L1 domain. Situating the current study within the previous SWORD studies opens the avenue for interpreting the meaning of the findings from the present study in connection with the larger body of SWORD studies. This connection between the current study and other SWORD studies may serve as a juncture where a host of findings and claims on students' peer feedback and revision activities meet to refine, refute and build on one another.

Second, this study offers a new method of conducting peer review in L2 writing classrooms. So far, the studies that examined peer review/feedback and writing of L2 writers

attend to the benefits of peer review and feedback activity in peripheral matters that are not directly related to writing development, such as developing writers' self-autonomy, improving critical reading skills, facilitating target language production, and enhancing audience awareness. But the current study revealed that the L2 writers produce more feedback comments than the L1 students at similar quality. This finding lays a theoretical ground for directing the future studies to further examine peer feedback as a viable source of feedback to students' writing in L2 setting. This in turn will increase opportunity to students to write more as part of school curriculum by allowing an alternative source of feedback to students' writing in addition to teacher feedback, which is expected to bring improvement in students' writing competency.

In addition, this study found that the improvement in L2 writers' essays are positively related to the percentage of criticism feedback incorporated during the revision, as well as to the amount of Type 4 and total revisions made during the revision process. This finding sheds light to the importance of feedback and revision in development of writing ability and sets a theoretical base to call for future studies to further investigate the topic.

Third, the present study is one of few that looked into the longitudinal effect of peer review and revision activity. The findings from this study suggest that the growth in writing quality gained from the revision process between the drafts may extend beyond the current drafts and add to the writers' general writing ability. It appears that the compositional benefits of revising a paper accumulate over time and are manifested through improvement in writing quality only when it reaches a certain threshold. Although this discussion is based on a small sample of students and is only in its inception, the present study is one of the first to introduce such findings, and may direct future studies to further investigate the topic.

5.2.2 Methodological Implications

This study draws two methodological implications. First, the coding scheme used in the present study to categorize students peer feedback comments is one of few reported in the field that is sophisticated enough to capture the meaningful features of feedback yet simple enough to be practical. The coding categories were based on the coding scheme developed

and used in the study by Patchen et al. (2009). The categories were simplified by deleting the categories found to be less revealing about the characteristics of peer feedback comments, such as the sub-categories of mitigation. In particular, the middle-level categories used to compare the characteristics of the feedback comments of L1 and L2 students in Chapter 4.1 consisting of praise, criticism, reflection; problem, solution, problem+solution; macro vs. micro; and meaning vs. surface were designed to allow analysis of both the quantitative and the qualitative characteristics of feedback comments. These categories delineate the characteristics of feedback comments from helpfulness perspective following the findings reported in the study that empirically tested helpfulness features of feedback comments (Nelson et al., 2009). The use of this coding scheme in future studies will allow researchers to compare and contrast feedback comments from functional (problem vs. solution), and helpfulness (specific vs. general; macro vs. micro) perspective.

Second, this study categorizes changes in the text into five types of revision, from Type 1 to Type 5, extending and updating the typology of revision developed by Faigley & Witte (1981). Most feedback studies, especially in L2 context, still use the dichotomous coding scheme of macro versus micro (or meaning versus surface) developed three decades ago. The extended typology of revision devised and first used by the researcher in this study encompasses more sophisticated classifications of revision types including the newly introduced Type 5 and allows investigation of the different functions of various types of revisions and their effects on the quality of paper with increased precision. The extended typology includes sub-divided the existing categories into four categories of punctuation/grammar, style/paraphrasing, extension/reduction of idea(s), and addition/deletion of idea(s). In addition, a type of revision that became more salient with the use of computer in writing was newly identified and added to the typology: re-organization of texts. By using this extended revision typology, the present study unveiled the importance of the Type 4 revision in improving the quality of student writing. The new revision typology system pioneered in this study will enable future studies to delineate the types of revision with enhanced sophistication.

5.2.3 Pedagogical Implications

In light of the findings, this study proposes five recommendations regarding implementation of peer feedback and revision sessions in classrooms. First, when employing peer feedback and revision to teach writing, multiple sessions of peer feedback and revision must be utilized, rather than a single session. Although the quality of students' feedback comments in this study did not improve over time with more peer feedback sessions, the students reported gaining growth in identifying issues in peers' writing, and devising solutions to problems in writing after multiple peer review sessions. In addition, engaging students in multiple sessions of peer review was found to have benefits beyond receiving and providing more and better feedback comments. This study found that students incorporated higher percentage of criticism feedback and made higher quality revisions as they moved from S1 to S2, suggesting that the students developed the ability to utilize feedback and make more meaningful revisions as they engage in multiple peer feedback and revision sessions.

Second, this paper recommends the instructors implementing peer feedback and revision sessions in class to provide some form of reward or control for the peer feedback and revision work of the students to ensure students' motivation and interest across sessions. For instance, the quantity and/or quality of peer feedback comments students generate should be evaluated (possibly through a back-evaluation from the writers) and counted towards the grading system. This study found that when the students' performance on the peer feedback and revision is not rewarded, i.e., not counted toward the course grade, students' performance on peer review activity generally declined in quantity and quality over time. The need to include peer review and revision in a grading system is corroborated by a student survey, in which a majority of students reported that they would behave differently in engaging in peer review activity if it were counted toward the course grade. It needs to be noted that simply engaging students in more peer feedback sessions did not bring improvement in the quality of students' peer feedback comments, even though it helped students' understanding and implementation of peer feedback comments. While it remains to be investigated in future studies what are the benefits of maintaining students' motivation level throughout the

repeated peer feedback sessions, it is expected to go beyond the benefits of repeated peer feedback activities found from the present study where students' motivation level declined.

In addition, the study offers implications for teacher training. The study showed that the instructors, even the ones who are supportive of peer review and feedback as the ones who participated in the study, did not fully utilize SWORD. They were not comfortable using peer rating and feedback as part of students' course grades and did not reward/penalize good/bad feedback. This reflects their uncertainty for the effects and the benefits of the peer review and feedback on writing development of students, which is largely due to lack of knowledge on the findings from peer feedback studies. Training programs especially for writing teachers are advised to address peer feedback more extensively to enable teachers to utilize this useful method for writing instruction in ideal conditions that may bring out the best results.

Third, composition instructors are advised to pay more attention to the students' revision processes and assist them in developing skills needed to make Type 4 revisions. This study found that employing Type 4 revisions leads to improved writing quality. Treglia (2009) claims that students who fail to address feedback often do so because of the difficulty involved in carrying out the challenging analytical tasks, and calls for teachers to provide specific guidance to help students achieve these challenging tasks. According to McCutchen (2000), writing expertise depends on the development of 1) fluent language generation processes and 2) extensive knowledge relevant to writing (e.g., genre knowledge). As such, students who have difficulty making Type 4 revisions may be helped by enforcing their language generation processes through activities such as sentence combining activities and/or by improving their knowledge on the specific genre.

Fourth, the instructors may benefit from knowing the specific characteristics pertaining to L1 and L2 writers' peer feedback comments, particularly when instructing classes with a mix of L1 and L2 population. For example, the present study found that the L1 students attend more to the solutions and L2 learners to the problems when providing feedback. It would help the students if the instructor discusses these points with the students

prior to peer feedback and try to re-balance their orientations so that both groups of students generate comments that contain balanced amount of problem and solution comments. Also, this study found that the feedback comments of L1 and L2 students display differences on certain text features, such as the use of mitigation language when providing criticism and the use of various modal verbs. These features may be explained by pragmatic aspect of language learning using Face-Threatening Acts and politeness strategies and need to be explicitly taught in writing classrooms comprised of a student body consisting of various linguistic/cultural background and interlanguage stages. As these differences may cause misunderstanding, offensiveness, or ineffective communication, instructor's discussion of these aspects with the students prior to peer review sessions will increase the effectiveness of the activity.

5.3 LIMITATIONS

The present study has a few limitations. First, the sample size used in the quantitative analyses of the study for research question one and two is relatively small compared to other SWORD studies. Although the total number of feedback segments analyzed (4,875 idea units) is larger than other SWORD studies, the number of participants were smaller than other SWORD studies as the students who participated in at least one peer review session ($N = 59$), and the number of students who participated in all three peer review sessions ($N = 32$) were reduced from the number of total participants of the study ($N = 66$), potentially affecting the generalization of the study's findings.

In addition, the analyses pertaining to research question three were conducted based on the data of small groups of cases due to the exploratory nature of the study. The number of samples in the improve groups was four ($N = 4$) each, and the number of participants in the non-improve group was smaller ($N = 3$ for L2 group and $N = 2$ for L1 group). This small sample size must be considered when interpreting the findings of the study and any attempt to generalize the findings should be made with caution.

Second, while the coding of peer feedback comments were tested for the inter-rater reliability and achieved a high agreement between the raters, the coding and counting of

revisions and incorporation of feedback in research question three was conducted solely by the researcher and was not tested for inter-rater reliability.

5.4 FUTURE RESEARCH

As the current study was one of the first attempts to compare the feedback comments of L1 and L2 learners and explore the longitudinal aspect of peer feedback and revision, more empirical endeavors are called for to verify findings from the study. This study proposes five viable research avenues for further investigation of peer feedback and writing.

First, the students were not rewarded for the quality/quantity of feedback comments they generated or for the revisions made between the drafts in this study, which apparently affected how they engaged in peer feedback and revision in long term and influenced the results on the longitudinal development of feedback comments. Future studies should investigate the changes/development in the peer feedback comments brought about by repeated engagement in peer review sessions where the performance of peer feedback is counted toward the course grade. This may reveal new findings about the developmental aspect of students' peer feedback comments. In addition, the prompts used for peer review in the current study remained the same across all three peer review sessions, despite the different genres of writing required of each session. It would be interesting to find out if students' interest and motivation for peer feedback sessions can be held up if the prompts were to be tailored to specifically suit each writing task.

Second, this study found that gains on students' writing brought by revising the drafts are not confined to improving the quality of the current paper but extends over to other writings based on the fact that the quality of draft 1 improved from S1 to S2 to S3 in the improve groups. This finding, however, was drawn from a small sample of students using constant comparison method. In analyzing the data, the quality of draft 1s of S1, S2, and S3

was not graded but derived from a conceptual formula. Future studies need to validate this finding by using a larger sample size and by using actual scores than the conceptual scores on draft 1.

Third, this was the first study that used SWORD in L2 classroom. Future studies can try using SWORD with L2 learners of lower language proficiency, or L2 learners in the EFL setting to see if SWORD can be successfully used by different groups of writers in different settings from the current and earlier studies. If SWORD can be used successfully in those settings, it will expand the boundary of SWORD studies and contribute to building a broader body of knowledge on peer review and writing.

Lastly, the findings on different textual features of feedback comments written by the L1 and the L2 groups can be further pursued into several follow-up studies. The feedback comments of L2 group can be analyzed by the students' first language to see if the usage of modals and mitigation vary based on students' linguistic/cultural background within the L2 group. Also, the same can be done based on students' years of stay or language proficiency to investigate its effect on development of pragmatics. Such analyses would help tease out the effects between the first language/culture versus the stage of linguistic development on second language learners' pragmatic development and allow more comprehensive account of such phenomena.

Appendices

APPENDIX A. STUDENT BACKGROUND SURVEY

If you are willing to participate in the study described above, please provide your name below.

Background Survey

1. What is your gender? _____ male _____ female
2. What is your age? _____
3. What is your major?

_ humanities _ social science _ business _ science/engineering _ undeclared
4. 4.1 What was your SAT verbal(critical reasoning) score (please use a number between 200 and 800-if you did not take the SATs, please type N/A)? _____

4.2 What was your SAT writing score (please use a number between 200 and 800-if you did not take the SATs, please type N/A)? _____

4.3 What was your SAT writing essay score (please use a number between 200 and 800-if you did not take the SATs, please type N/A)? _____
5. 5.1 What is your first language? _____

5.2 How long have you lived in the US(or English speaking country)?
_ less than 1 yr _ 1-3 yrs _ 4-7 yrs _ 8-10 yrs _ more than 10 yrs
6. 6.1 What was your total TOEFL(ibt) score (please use a number between 0 and 120-if you did not take TOEFL in past three years, please type N/A)? _____

6.2 What was your TOEFL(ibt) writing score (please use a number between 0 and 30-if you did not take the TOEFL in past three years, please type N/A)? _____

7. Here are some questions about your perception about writing in general. On a scale of 1(no chance) to 6 (completely certain), please indicate the rating that best describes what you think.

7.1 Good teachers can help me become a better writer.

1 2 3 4 5 6

7.2 Good writers are born, not made.

1 2 3 4 5 6

7.3 Some people have said, "Writing can be learned, but it can't be taught."

Do you believe that it can be learned?

1 2 3 4 5 6

7.4 Some people have said, "Writing can be learned, but it can't be taught."

Do you believe it can be taught?

1 2 3 4 5 6

7.5 I avoid writing (in first language or in general).

1 2 3 4 5 6

7.6 I avoid writing in second language.

1 2 3 4 5 6

7.7 I enjoy writing (in first language or in general).

1 2 3 4 5 6

7.8 I enjoy writing in second language.

1 2 3 4 5 6

7.9 Discussing my writing with others is an enjoyable experience.

1 2 3 4 5 6

7.10 I'm no good at writing (in first language or in general).

1 2 3 4 5 6

7.11 I'm no good at writing in second language.

1 2 3 4 5 6

7.12 I am a good writer (in first language or in general).

1 2 3 4 5 6

7.13 I am a good writer (in second language).

1 2 3 4 5 6

7.14 I believe I was born with the ability to write well.

1 2 3 4 5 6

8. Here are some questions about yourself as a student in this class. Please indicate the rating that best describes what you think.

8.1 It is important to me that other students in my class think I am good at my class work.

1 2 3 4 5 6

8.2 It is important to me that I learn new concepts and master the content of this class.

1 2 3 4 5 6

8.3 It is important to me that I get a good grade in this class.

1 2 3 4 5 6

8.4 It is important to me that I improve my English in this class.

1 2 3 4 5 6

9 9.1 Writing in first language is different from writing in second language by nature.

1 2 3 4 5 6

9.2 Which of the factors below would make you become a better writer when you improve on them? Please rank these from 1-4, where 1 is the most important and 4 the least.

___language proficiency (English)

___ idea generation & reasoning (thinking)

___ understanding of (American) culture ___ understanding of genres/rhetorical features

10.1 How many times have you engaged in a peer-review of writing before this class?

___ none ___ 1-2 ___ 3-4 ___ 5-7 ___ 8 or more

10.2 In your previous experiences with peer review, what kind of (if any) have you had?

I was given a clear explanation of purpose and goals of peer review.

1 2 3 4 5 6

I was given a form with questions to answer.

1 2 3 4 5 6

I was given examples of graded papers.

1 2 3 4 5 6

I was given examples of completed score sheets.

1 2 3 4 5 6

APPENDIX B. STUDENTS' POST SEMESTER SURVEY

Course: _____ Name: _____

Peer Review Experience

Over the semester, you had a chance to engage in three sessions of **peer review** activities in this course. Please take a moment to think about your experience in each of these sessions. These questions pertain ONLY to **PEER review** and NOT to teacher review.

1. Do you think the types, content, and/or quality of feedback that you received and gave changed over the semester? Yes _____ No _____

2. Here are some of the types/contents that describe the feedback you **received** from peers. On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what you think.

2.1 Over the semester, I became more interested in seeing my peers' feedback on my paper.

1 2 3 4 5 6

2.2 Over the semester, I accepted/ incorporated more peer feedbacks in my revision.

1 2 3 4 5 6

2.3 Over the semester, I came to make better sense out of/understand better the feedback I received.

1 2 3 4 5 6

2.4 Over the semester, the feedback I received became more helpful in my revision and writing.

1 2 3 4 5 6

3. Here are some of the types/contents that describe the feedback you **gave** to your peers. On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what you think.

3.1 Over the semester, I became more interested and engaged in providing feedback.

1 2 3 4 5 6

3.2 Over the semester, I spent more time and effort in providing feedback to my peers.

1 2 3 4 5 6

3.3 Over the semester, it became easier to identify problems in my peers' paper.

1 2 3 4 5 6

3.4 Over the semester, it became easier to come up with solutions to the problems identified.

1 2 3 4 5 6

3.5 Over the semester, I think my feedback became more accurate.

1 2 3 4 5 6

3.6 Over the semester, I think my feedback became more helpful to peers in their revision and writing.

1 2 3 4 5 6

3.7 Over the semester, if you spent less time and effort in providing feedback to your peers, what would be the reason?

1. Had other assignments that were more important to me
2. Peer review became less interesting and engaging

3. Found less value in peer review (my feedback didn't seem to be valued or helping others)
4. Peer review became easier and thus took less time and effort
5. Other (please specify:

3.8 Over the semester, if you spent more time and effort in providing feedback to your peers, what would be the reason?

1. Peer review became more important for my success/grade in the course.
2. Peer review became more interesting and engaging.
3. I found my comments to be valued and helping my peers
4. Peer review became more difficult requiring me to spend more time and effort.
5. Other (please specify:

3.9 If your peer review comments were reviewed and graded as part of the course grade, would you have done it differently?

Yes _____ No _____

If yes, what would be the major difference?

1. I would have read my peers' paper more thoroughly to provide accurate feedback.
2. I would have tried to provide more detailed and specific comments.
3. I would have been more sensitive to deadlines and participate in all sessions.
4. Other (please specify:

4. Think about the comments you received and how you incorporated or rejected certain suggestions. On a scale of 1 (strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what affected/will affect your acceptance/rejection of a feedback.

4.1 I incorporate feedback in my revision when I agree with them.

- | | | | | | | |
|--|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.2 I incorporate feedback that I don't necessarily agree with when I get the same feedback from multiple peers. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.3 I incorporate feedback that provides solutions to the problem. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.4 I incorporate feedback which is detailed, specific and easy to understand. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.5 I incorporate feedback that seems to have taken much time and effort on the part of the feedback provider. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.6 I reject feedback that requires significant rework on my writing. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.7 I reject feedback that is vague and abstract. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.8 I reject feedbacks that are degrading regardless of their content. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 4.9 I think criticism is more helpful than praise as a feedback. | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
5. Please indicate the approximate percentage of peer feedback that you incorporated in your revision on a scale of 0(none) to 100% (all) in your revision for each peer review sessions over the semester.

5.1 First peer review session (on the first writing assignment) _____ %

5.2 Second peer review session (on the second writing assignment) _____ %

5.3 Third peer review session (on the third writing assignment) _____ %

If the percentage changed across sessions, what would be the reason(s) for the change?

6. On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes the goal you had when you provided feedback to your peers.

6.1 When providing peer feedback, my primary goal was to complete the task in the shortest time possible.

1 2 3 4 5 6

6.2 When providing peer feedback, my primary goal was to fill up the comment space provided.

1 2 3 4 5 6

6.3 When providing peer feedback, my primary goal was to identify problems of my peers' writing and to provide helpful solutions.

1 2 3 4 5 6

6.4 When providing peer feedback, my primary goal was to improve my critical reading skills.

1 2 3 4 5 6

7 The following questions ask your perception in general about the feedback you gave and received during this course. On a scale of 1 (strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what you think.

7.1 Feedback I gave my peers on their writing was/will be useful to their writing.

1 2 3 4 5 6

7.2 Feedback I gave my peers would be similar to the feedback that other peers or teacher gave/will give on the same paper.

1 2 3 4 5 6

7.3 Feedback my peers gave me on my writing was/will likely to be useful.

1 2 3 4 5 6

7.4 If I had to give feedback several months from now on the same papers for which I gave feedback in this class, I would probably give similar feedback.

1 2 3 4 5 6
7.5 Feedback I got from one peer was similar to the feedback I got from other peers on the same paper.

1 2 3 4 5 6
7.6 If my peers gave me feedback several months from now on the same paper they examined for this class, they would probably give me similar feedback.

1 2 3 4 5 6
7.7 Peers gave me a fair rating on my writing.

1 2 3 4 5 6
7.8 Reviewing four papers (as opposed to one or two) helped me become better at identifying problems in and providing solutions for papers.

1 2 3 4 5 6
7.9 Receiving feedback from multiple peers helped me revise my paper better.

1 2 3 4 5 6
7.10 Doing peer review three times over the semester (as opposed to doing it just once) has helped me become better at identifying problems in and providing solutions for writing.

1 2 3 4 5 6
7.11 Doing peer review three times over the semester (as opposed to doing it just once) has helped me become better at revising my papers.

1 2 3 4 5 6
Writing Experience

8 For each of your three writing assignment, you wrote three drafts. On a scale of 1(minimum) to 6(maximum), please indicate the RELATIVE time and effort you put in for each draft.

8.1 First assignment draft 1 (version 1.1, before peer review)

1 2 3 4 5 6

8.2 First assignment draft 2 (version 1.2 after peer review)

1 2 3 4 5 6

8.3 First assignment draft 3 (version 1.3 after teacher review)

1 2 3 4 5 6

8.4 Second assignment draft 1 (version 2.1, before peer review)

1 2 3 4 5 6

8.5 Second assignment draft 2 (version 2.2 after peer review)

1 2 3 4 5 6

8.6 Second assignment draft 3 (version 2.3 after teacher review)

1 2 3 4 5 6

8.7 Third assignment draft 1 (version 3.1, before peer review)

1 2 3 4 5 6

8.8 Third assignment draft 2 (version 3.2 after peer review)

1 2 3 4 5 6

8.9 Third assignment draft 3 (version 3.3 after teacher review)

1 2 3 4 5 6

9 On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes your writing behavior and changes you had over the semester.

9.1 Over the semester, I think I became better at identifying problems in my own writing.

1 2 3 4 5 6

9.2 Over the semester, I think I became better at coming up with solutions to the problems I find in my writing.

1 2 3 4 5 6

9.3 Over the semester, I came to have a better understanding about what makes a good writing.

1 2 3 4 5 6

9.4 Over the semester, I think I became better at telling a better writing from a worse writing when I see one.

1 2 3 4 5 6

9.5 Over the semester, I think I came to write with more ease.

1 2 3 4 5 6

9.6 Over the semester, I think my writing improved in the areas of...

Clearly stating my thesis/argument and lining up the whole paper around it.

1 2 3 4 5 6

Generating/selecting the contents/details for the paper to make the thesis strong and persuasive.

1 2 3 4 5 6

Organizing the paper to have logical and smooth flow.

1 2 3 4 5 6

Using more sophisticated and diverse sentence structure and vocabulary.

1 2 3 4 5 6

Others(please describe)

9.7 When I worked on my second and third assignments, I was reminded of some of the feedbacks I received from peers on my earlier assignment(s).

1 2 3 4 5 6

9.8 When I worked on my second and third assignments, I was reminded of some of the feedbacks I gave to my peers in earlier assignment(s).

1 2 3 4 5 6

SWoRD Experience

Over the semester, you used online peer review tool named SWoRD in your peer review activity. The following questions are about your experience in using SWoRD. On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what you think.

10.1 Using SWoRD made peer review more accessible and easy.

1 2 3 4 5 6

10.2 Using SWoRD made peer review more complicated and taxing.

1 2 3 4 5 6

10.3 Using SWoRD improved the quality of my writing.

1 2 3 4 5 6

10.4 Using SWoRD improved the quality of my peer review.

1 2 3 4 5 6

10.5 Back evaluation comments I got on my feedback helped me give more helpful feedback to my peers.

1 2 3 4 5 6

10.6 I referenced/tried to keep in mind the back evaluation comments I got on my feedback when providing feedback to my peers in the next sessions.

1 2 3 4 5 6

10.7 I got a fair rating through peer review.

1 2 3 4 5 6

10.8 I think it is reasonable to use peer ratings for assigning grades for papers.

1 2 3 4 5 6

10.9 I think this course should continue using SWoRD.

1 2 3 4 5 6

10.10 I think other courses should adopt using SWoRD.

1 2 3 4 5 6

In case the researcher has further questions regarding your peer review and writing experiences in this class, would it be okay for the researcher to contact you during the next few months?

____ Yes, the researcher may contact me via my email_____.

____ No, I do not want to be contacted by the researcher.

APPENDIX C. INSTRUCTOR SURVEY

Course: _____

Peer Review

Over the semester, your students had a chance to engage in three sessions of **peer review** in this class. Please take a moment to recall your instructions in class, observations and insights relating to peer review sessions.

1. Over the semester, I discussed about peer review –its purpose, benefits, how to provide a good peer review etc- in class approximately _____ times/sessions totaling in _____ minutes.
2. Please describe the type of materials, thoughts, talks on peer review that you shared with your students before each peer review session (if none, put none) and approximate time you spent on it in class.

Session 1:

Session 2:

Session 3:

Other time:

3. Please describe the settings of your students peer review activity for each session over the semester. (in-class vs. homework, how much time was given,)

Session 1:

Session 2:

Session 3:

4. For the peer review sessions that took place in class (either in part or full), please rate the degree of students' engagement on a scale of 1(least) to 6(most) as was observed by you for each session.

Session 1	1	2	3	4	5	6
Session 2	1	2	3	4	5	6
Session 3	1	2	3	4	5	6

5. I believe in the benefits of peer review and am a strong supporter of it.

1 2 3 4 5 6

6. I think the primary benefit/purpose of peer review to students is(are):

_____.

7. I think that students are capable of providing feedbacks that are valid (accurate).

1 2 3 4 5 6

Writing

8. Over the semester, you reviewed total six drafts (two drafts for three assignments) of writing of each of your student. Please indicate the degree of your agreement with the following statements on a scale of 1(strongly disagree) to 6 (strongly agree).

8.1 I see a clear improvement, in general, between version 1(draft) and version 2(final) of the assignments.

1 2 3 4 5 6

8.2 I see a clear incremental improvement, in general, between version 1 of the first assignment to version 1 of the second assignment, to the version 1 of the third assignment.

1 2 3 4 5 6

8.3 I see a clear incremental improvement across version 2 of the first to second to third assignments.

1 2 3 4 5 6

8.4 Please list top five topics in your curriculum that you spent most time discussing in class and believe to be most important for students' success in their writing improvement. (please indicate approximate total time spent on each topic)

1.

2.

3.

4.

5.

9. Over the semester, I think my students' writing improved in the areas of...

Clearly stating my thesis/argument and lining up the whole paper around it.

1 2 3 4 5 6

Generating/selecting the contents/details for the paper to make the thesis
strong and persuasive.

1 2 3 4 5 6

Organizing the paper to have logical and smooth flow.

1 2 3 4 5 6

Using more sophisticated and diverse sentence structure and vocabulary.

1 2 3 4 5 6

Others

10. Please indicate the approximate percentage of your feedback that you think was incorporated in your students revision on a scale of 0(none) to 100% (all) for each assignment over the semester.

10.1 First writing assignment _____ %

10.2 Second writing assignment _____ %

10.3 Third writing assignment _____ %

If the percentage changed across sessions, what do you think is the reason for the change? _____

11. What do you see as the area(s) that needs to be improved the most in your students' writing? _____

12. Based on your past experience of teaching similar courses, are the group of students you had for this class more or less the typical group in terms of their attitude, behavior and engagement in class assignments and activities?
(if not, please describe how they are different)

SWoRD

Over the semester, you used online peer review tool named SWoRD in your class peer review activity. The following questions are about your experience in using SWoRD. On a scale of 1(strongly disagree) to 6 (strongly agree), please indicate the rating that best describes what you think.

13.1 Using SWoRD made peer review more accessible and easy.

1 2 3 4 5 6

13.2 Using SWoRD made peer review more complicated and taxing.

1 2 3 4 5 6

13.3. Using SWoRD improved the quality of my students' writing.

1 2 3 4 5 6

13.4 Using SWoRD improved the quality of students' peer review feedback.

1 2 3 4 5 6

13.5 I think it is reasonable to use peer ratings for assigning grades for papers.

1 2 3 4 5 6

13.6 I plan to continue to use SWoRD for this course.

1 2 3 4 5 6

13.7 I think other classes/courses should adopt using SWoRD.

1 2 3 4 5 6

Any other comments regarding SWoRD?

Thank you for your participation!!!

Examples of Good and Bad Feedback

From actual students' feedback

Avoid whole-sale comments:

explain why you think something is good or bad

"This paper has a clear thesis, a well structured introduction and conclusion."

"I had a hard time pin pointing your thesis."

"Good arguments, interesting topic!"

"Clarify thesis."

- Writer says, *"No advice was given. They mention that introduction was good, but no real feedback.."*

Be specific!

“Grammar and punctuation errors, parenthesis errors.”

“The paper was well written and flowed well but some of the points could have been clearer.”

Writer says, *“when you don’t describe what is unclear, how can I make things clearer?”*

Avoid Insulting Comments

“ It’s a solid idea and you make some pretty good insights, but nothing really new. This is all stuff I’ve been hearing since I was in seventh grade”

“Your statement “~~~” is wrong. If you got that from a source, double check it. And if you just pulled it out of your head, try to research it a little.”

Writer says, *“You are condescending”*

Avoid giving only and all praises

“Your paper flowed smoothly from one topic to the next.”

“The logic of the arguments was good and points that were used did support the main topic of this paper.”

“Good use of sources and good organization.”

“Good topic sentences. Effective transitions.”

A writer says, *“so I have nothing to fix?”*

Another writer says, *“Had I only seen your comments, I would’ve thought my paper was perfect!”*

Some good feedback:

identify problem, explain reason and provide solution

“I think that if attention is your main argument, the beginning of your paper about proper language vs. vernacular and the part about movies being influential are not needed at all. I can’t find how these relate to the rest of your paper. So I would recommend either getting rid of them or clarifying how they relate to your main argument.”

Writer says, *“very helpful feedback. I felt that you gave the paper a close reading, which I appreciated. Both your positive and negative comments were helpful because they were specific. Thanks”*

Feedback providing examples

“The second paragraph can be broken up into two paragraphs. You can break it up at the sentence *“We may even compare...according to which approach they take.”* because you seem to be switching ideas.”

“ Also some sentences tend to be lengthy and a little confusing. Like in paragraph 4, the sentence *“The rush of lyrics...leads to the extinction of whales”*, is lengthy and a little hard to follow what you are trying to say.”

Feedback with specific solutions

“The title of the article needs to be in quotations; the author does not have to be italicized.”

“I noticed some paragraphs hang at the end. Try to make sentences at the end to either summarize or pinpoint the importance of that paragraph! “

- Writer says, *“Thanks for the advice. This is very helpful because now I know what I need to do to fix the problem”*

Excerpts from comparison of Wendy's draft 1 and draft 2 in S2

152

because she speaks to people of the U.S. who are affected by the ~~is in a~~ recession and are trying their best to save, and informs them that investing would cost them ~~saving money means~~ a lot of money. She also argues that ~~does this by stating~~ "Batteries aren't the most durable materials on earth; they go out pretty fast, then what, More money?" (3). ~~By taking this different approach of logos, she explains (3)?~~ She then takes a logos-reasoning appeal by stating, while you won't need to purchase fuel or pay for expensive maintenance associated with traditional vehicles, the cost of the battery for the electric car could eat up the savings. She uses the rhetorical strategy of a slippery slope by saying if you invest in electric cars you won't have any money left. (This is an effective way of stating her belief that electric cars aren't as good an investment as perceived.)

Comment [u13]: 13) Elaboration: Type 3

Comment [u14]: 14) paraphrase: Type 2

Comment [u15]: 15) key term added: Type 3

Comment [u16]: 16) summary added: Type 3

(Then she uses pathos in order to argue against the efficiency of electric cars. She talks about the ~~then moves on to~~ driving distance for electric vehicles which typically only make it between 100 and 150 miles, depending on the vehicle, before needing a recharge. Here she appeals to the current or potential electric car drivers emotions of anger by presenting a shocking statistic.) This appeal is effective as ~~and how~~ most consumers are accustomed to driving a vehicle for several hundred miles before ~~refueling, and with~~ America being a demanding country, this would persuade people to side with the author. She goes on to say, when refueling. When the tank nears the empty mark, you ~~make a~~ "simple stop" at a gas station and refuel. She uses words such as "simple ~~simple stops~~ steps" to emphasize the easiness of the traditional gas stop, when comparing to "long 8 hour recharges". She ends the paragraph by stating, for stop. ~~She goes on to say~~ electric vehicles typically only make it between 100 and 150 miles, depending on the vehicle, before needing a recharge. For every day short distance commutes, the limited driving range doesn't present a problem, but longer distance driving, such as long commutes or road trips, "becoming ~~become~~ more of a challenge if you need to stop every 100 miles to recharge the battery of your vehicle" (4). Once again she relates the negatives to American drivers, as she knows that Americans like to travel and take vacations often, she uses "Knowing only 100 miles" as she knows that this insufficient number of miles wouldn't be enough mileage in which the electric car can go without a charge, further strengthening ~~for the average~~ American driver she uses this to strengthen her argument.

Comment [u17]: 17) topic sentence added: Type 3

Comment [u18]: 18) Paraphrase: Type 2

Comment [u19]: 19) sentence moved up from below: Type 5

Comment [u20]: 20) elaboration: Type 3

Comment [u21]: 21) paraphrase: Type 2

Comment [u22]: 22) Specifics added: Type 3

Comment [u23]: 23) paraphrase: Type 2

Comment [u24]: 19) sentence re-located: Type 5

Comment [u25]: 24) summary added: Type 3

Comment [u26]: 25) paraphrase: Type 2

References

- Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Ash, B. H. (1983). Selected effects of elapsed time and grade level on the revisions in eighth, tenth, and twelfth graders' writing. *Dissertation Abstracts*, 43 (12), 3830A.
- Bamberg, B. (1978). Composition instruction does make a difference: A comparison of college freshmen in regular and remedial English courses. *Research in the Teaching of English*, 12, 47-59.
- Bartlett, E. (1981). *Learning to Write: Some Cognitive and Linguistic Components*. Washington, DC: Center for Applied Linguistics.
- Beach, R. (1976). Self-evaluation strategies of extensive revisers. *College Composition and Communication*, 27, 160-164.
- Beason, L. (1993). Feedback and Revision in Writing Across the Curriculum Classes. *Research in the Teaching of English*, 27 (4), 395-422.
- Benjamin, R., & Chun, M. (2003). A new field of dreams: The Collegiate Learning Assessment Project. *Peer Review*, 5(4), 26-29.
- Benson, B. (1980). A qualitative analysis of language structures in compositions written by first and second language learners. *Dissertation Abstracts International*, 41(5), 1984A.
- Benson, B., Deming, M., Denzer, D., & Valeri-Gold, M. (1992). A combined basic writing/English as a second language class: Melting pot or mishmash? *Journal of Basic Writing*, 11(1), 58-59.
- Berg, B. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215-241.
- Berlin, J. (1987). *Rhetoric and reality: Writing instruction in American colleges, 1900-1985*. Carbondale: Southern Illinois University Press.
- Bond, S., Hayes, J., Flower, L. (1980). *Translating the Law into Common Language: A Protocol Study*. Document Design Project 1, Pittsburgh, PA: Carnegie-Mellon University.
- Bracewell, R. J., Scardamalia, M., & Bereiter, C. (1978). *The development of audience awareness in writing*. (ERIC Document Reproduction Service No. ED 154 433)
- Bridwell, L. S. (1980). Revising strategies in twelfth grade students' transactional writing. *Research in the Teaching of English*, 14, 197-222.
- Bridwell, L., Sire, G., & Brooke, R. (1985). Revising and computing.

- Brown, P., & Levinson, S. (1978). Politeness: some universals in language usage. In Goody, E. N. (ed.) *Questions and politeness*. Cambridge: Cambridge University Press. 56-310.
- Brown, P., & Levinson, S. (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Bruffee, K. (1978). The Brooklyn Plan: Attaining intellectual growth through peer-group tutoring. *Liberal Education*, 64, 447-468.
- Bruffee, K. (1984). Collaborative learning and the conversation of mankind. *College English*, 28 (1), 181-188.
- Bruffee, K. (1993). *Collaborative learning: Higher education, interdependence and the authority of knowledge*. Baltimore: Johns Hopkins University Press.
- Butler, J., Britt, M. (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication*, 28(1), 70-96.
- Buxton, E. W. (1959). An experiment to test the effects of writing frequency and guided practice upon student mechanics skill in writing expression. *Alberta Journal of Educational Research*, 5, 91-99.
- Byrd, D. R. (1994). Peer editing: Common concerns and applications in the foreign language classroom. *Die Unterrichtspraxis/Teaching German*, 21(1), 119-123.
- Campbell, C. (1987a). Writing with others' words: Native and non-native university students use of information from a background reading text in academic compositions. (ERIC Document Reproduction Service No. ED 287315)
- Campbell, C. (1987b). Writing with others' words: The use of information from a background reading text in the writing of native and non-native university composition students. *Dissertation Abstracts International*, 48(7), 1679A.
- Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 211-230). New York: Cambridge University Press.
- Carlson, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 227-260). Newbury Park, CA: Sage.
- Carson, J. G., Nelson, G. L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing*, 5, 1-19.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28, 181-188.
- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In: Chafe, W., Nichols, J. (Eds.), *Evidentiality: The Linguistic Coding of Epistemology*. Ablex, Norwood, NJ, pp. 261-272.

- Channell, J. (1994). *Vague Language*. Oxford University Press, Oxford.
- Charney, D., Newman, J., Palmquist, M. (1995). "I'm Just No Good at Writing" Epistemological Style and Attitudes Toward Writing. *Written Communications*, 12(3), 298-329.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, 19, 151-165.
- Chein, I. (1981). Appendix: An introduction to sampling. In Kidder, L.H. (ed), *Selltiz, Wrightman & Cook's Research Methods in Social Relations*. (4th ed) Austin Tx.: Holt, Reinehart and Winston.
- Cho, Y. & Cho, K. (2010). Peer reviewers learn from giving comments. *Instructional Science*, 39, 629-643.
- Cho, K., & Schunn, C. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, 48 (3), 409-426.
- Cho, K., Schunn, C., & Charney, D. (2006). Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts. *Written Communication*, 23 (3), 260-294.
- Cho, K., Schunn, C., & Wilson, R. (2006). Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology*, 98 (4), 891-901.
- Chou, M. C. (1999). How peer negotiations shape revisions. In J. Katchen, & Y. N. Leung (Eds.), *The Proceedings of the Seventh International Symposium on English Teaching* (pp. 349–359). Taipei: The Crane Publishing Co.
- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics*, 17(1-4), 301-316.
- Cummings, V. (1990). Speech and writing: An analysis of expository texts compared by native and non-native speakers of English at The City University of New York. *Dissertation Abstracts International*, 51(7), 2296A.
- de Guerrero, M., & Villamil, O. (1994). Social cognitive dimensions of interaction in L2 peer revision. *The Modern Language Journal*, 78(4), 484–496.
- de Larios, J.R., Manchon, R.M., & Murphy, L. (2006). Generating text in native and foreign language writing: a temporal analysis of problem solving formulation processes, *Modern Language Journal*, 90 (1), 100-114.
- Elbow, P. (1973). *Writing Without Teachers*. New York: Oxford University Press.

- Emig, J. (1971). The composing processes of twelfth graders. *Research Report No.13*. Urbana, Ill.: National Council of Teachers of English.
- Ericsson, A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K.A. Ericsson, N.Charness, P.J Feltovich, & R. R Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.683-703). Cambridge UK: Cambridge University Press.
- Ericsson, A., Krampe, R.T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363-406.
- Ericsson, K. A., Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32, 400-414.
- Ferris, D. (1997). The Influence of Teacher Commentary on Student Revision. *TESOL Quarterly*, 31(2), 315-339.
- Flower, L., Hayes, J., Carey, L., Schriver, K., Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. *College Composition and Communication*. 37 (1), 16-55.
- Flynn, E (1982, March). Freedom, restraint, and peer group interaction. Paper presented at the Annual Meeting of the Conference on College Composition and Communication, San Francisco, CA.
- Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research*, 54, 363-407.
- Frodesen, J. (1991). Aspects of coherence in a writing assessment context: Linguistic and rhetorical features of native and non-native English essays. *Dissertation Abstracts International*, 52(1), 150A.
- Gere, A. R., & Stevens, R. S. (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.
- Geiser, S., Studley, R. (2001). *UC and the SAT: Predictive validity and differenetial impact of the SAT I and SAT II at the Univsersity of California*. Oakland: University of California Office of the President.
- Glaser, B.G., Strauss, A.L. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Graham, S. & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445-476.
- Graner, M. (1987). Revision Workshops: An alternative to peer editing groups. *The English Journal*, 76, 40-45.

- Hall, C. (1990). Managing the complexity of revising across languages. *TESOL Quarterly*, 24(1), 43-60.
- Hafernik, J. (1990). Relationships among English writing experience, contrastive rhetoric, and English expository prose of L1 and L2 college writers. *Dissertation Abstracts International*, 51(12), 4007A.
- Hansen, J., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal*, 59 (1), 31-38.
- Harris, M. (1990). Teacher/student talk: The collaborative conference. In S. Hynds,&D. Rubin (Eds.), *Perspective on talk and learning* (pp. 149–161). Urbana, IL: National Council of Teachers of English.
- Haswell, R. (2005). NCTE/CCCC's Recent War on Scholarship. *Written Communication*, 22 (2), 198-223.
- Hattie, J. A. (1999). *Influences on student learning* (Inaugural professorial address, University of Auckland, New Zealand). Retrieved from <http://www.arts.auckland.ac.nz/staff/index.cfm?P=8650>
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In M. Levy & S. Ransdell (Ed.), *The Science of Writing* (pp.1-27). Mahwah, NJ: Lawrence Erlbaum Association.
- Hayes, J., Flower, L., Schriver, K., Stratman, J., Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in psycholinguistics, volume II: Reading, writing, and language processing* (pp.176-240). Cambridge, England: Cambridge University Press.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77 (1), 81–112.
- Hillocks, G. (1982, October). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English*, 16, 261-278
- Hillocks, G. (1984). What Works in Teaching Composition: A Meta-analysis of Experimental Treatment Studies. *American Journal of Education*, 93 (1), 133-170.
- Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37 (2), 275-301.
- Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41, 667-683.
- Hyland, F., Hyland, K. (2001). Sugaring the pill: praise and criticism in written feedback. *Journal of Second Language Writing*, 10 (3), 185-212.

- Johnstone, K., Ashbaugh, H., & Warfield, T. (2002). Effects of repeated practice and contextual-writing experiences on college students' writing skills. *Journal of Educational Psychology*, 94 (2), 305-315.
- Kaufman, J., Schunn, C. (2010). Students' perception about peer assessment for writing: their origin and on revision work. *Instr Sci* DOI 10.1007/s11251-010-9133-6
- Keselman, H. J., Kowalchuk, R. K., & Lix, L. M. (1998). Robust non-orthogonal analyses revisited: An update based on trimmed means. *Psychometrika*, 63, 145-163.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Randell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp.75-71). Mahwah, NJ: Erlbaum.
- Kellogg, R., Whiteford, A. (2009). Training Advanced Writing Skills: The Case for Deliberate Practice. *Educational Psychologist*, 44(4), 250-266.
- Lee, I. (2011). Working Smarter, Not Working Harder: Revisiting Teacher Feedback in the L2 Writing Classroom. *The Canadian Modern Language Review*, 67 (3), 377-399.
- Leech, G. (2003). Modality on the move: the English modal auxiliaries 1961–1992. In: Facchinetti, R., Krug, M., Palmer, F. (Eds.), *Modality in Contemporary English*. Mouton de Gruyter, Berlin/New York, pp. 223–240.
- Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp.57-68). New York: Cambridge University Press.
- Li, L., Liu, X., & Steckelberg, A. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41 (3), 525-536.
- Lin, C. (1989). The structures of English and Chinese narratives written by college students in Taiwan. *Dissertation Abstracts International*, 50(7), 2036A.
- Liu, J., Hansen, J. (2002). *Peer Response in Second Language Writing Classrooms*. Ann Arbor: The University of Michigan Press. (ISBN 0-472-08808-4)
- Liu, J., Randall, S. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing, *Journal of English for Academic Purposes*, 2, 193-227.
- Lockhart, C., & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning*, 45, 605–655.
- Long, M., & Porter, P. (1985). Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly*, 19 (2), 305-325.
- Lowenthal, D. (1980). Mixing levels of revision. *Visible Language*, 14(4), 383-387.

- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18 (1), 30-43.
- Lynch, C., & Klemans, P. (1978). Evaluating our Evaluators. *College English*, 4 (2), 166-180.
- MacArthur, C., Graham, S., & Schwartz, S. (1991). Knowledge of revision and revising behavior among learning disabled students. *Learning Disability Quarterly*, 14 (1), 61-73.
- McCutchen, D. (2000). Knowledge, processing, and working memory: implications for a theory of writing. *Educational Psychology*, 35(1), 13-23.
- Maynor, L.C., (1982). An investigation of the revising practices of college freshman writers. *Dissertation Abstracts*, 43(8), 2543A.
- Markkanen, R., Schroeder, H. (1997). Hedging: a challenge for pragmatics and discourse analysis. In: Markkanen, R., Schroder, H. (Eds.), *Hedging and discourse: approaches to the analysis of a pragmatic phenomenon in academic texts*. Walter de Gruyter, Berlin, pp. 3-20.
- Mendonca, C., & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 4, 745-769.
- Merriam, S. (1998). *Qualitative research and case study application in education. Revised and expanded from case study research in education*. A joint publication in the Jossey-Bass Education Series and the Jossey-Bass Higher Education Series.
- Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher Feedback in Chinese EFL writing class. *Journal of Second Language Writing*, 15 (4), 179- 200.
- Min, H. T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293-308.
- Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15 (2), 118-141.
- 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: Longman.
- Monahan, B. D. (1982). *Revision strategies of basic and competent writers as they writer for different audiences*. (ERIC Document Reproduction Service No. ED 229 756)
- Murray, D. (1978a). Internal revision: A process of discovery. In C. Cooper and L. Odell (Eds.), *Research on composing: Points of departure* (pp. 85-103). Urbana, IL: National Council of Teachers of English.

- National Assessment of Educational Progress. (1977). *Write/rewrite: An assessment of revision skills: Selected results from the second national assessment of writing*. (ERIC Document Reproduction Service No. ED 141 826)
- National Assessment of Educational Progress. (2007). *The Nations Report Card. Writing 2007 Major Results*. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/writing_2007/
- National Commission on Writing. (2004, September). *Writing: A ticket to work . . . or a ticket out: A survey of business leaders*. Retrieved from http://www.writingcommission.org/prod_downloads/writingcom/writing-ticket-to-work.pdf
- Nelson, G., Murphy, J. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27 (1), 135-142.
- Nelson, M., Schunn, C. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 27(4), 375-401.
- Paulus, T. (1999). The Effect of Peer and Teacher Feedback on Student Writing. *Journal of Second Language Writing*, 8(3), 265-289.
- Patchan, M., Charney, D., & Schun, C. (2009). A validation study of students' end comments: Comparing comments by students, a writing instructor, and a content instructor. *Journal of Writing Research*, 1 (2), 124-152.
- Patchan, M. M., Schunn, C.D., & Russell, R. (2011). Writing in natural sciences: Understanding the effects of different types of reviewers on the writing process. *Journal of Writing Research*, 2(3), 365-393.
- Patton, M.Q. (1990). *Qualitative Evaluation Methods*. (2nd edition). Thousand Oaks, California: Sage.
- Park, Y. (1988). Academic and ethnic background as factors affecting writing performance. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 261-272). Newbury Park, CA: Sage.
- Perkins, M.R. (1983). *Modal Expressions in English*. Pinter, London.
- Perl, S. (1980, December). Understanding composing. *College Composition and Communication*, 31, 363-369.
- Peterson, S. (2003). Peer response and students' revisions of their narrative writing. *L1-Educational Studies in Language and Literature*, 3, 239-272.
- Radecki, P., & Swales, J. (1988). ESL Students' Reaction to Written Comments on Their Written Work. *System*, 16 (3), 355-365.
- Reid, J. (1988). Quantitative differences in English prose written by Arabic, Chinese, Spanish, and English students. *Dissertation Abstracts International*, 50(3), 672A.

- Rollinson, P. (1998). Peer response and revision in an ESL writing group: a case study. Unpublished PhD thesis. Universidad Autonoma de Madrid.
- Rubin, L. (1984). *An Investigation of Self-Evaluation: How Student Writers Judge Their Writing Process and Product*. Dissertation. Carnegie-Mellon University.
- Saito, H., & Fujita, T. (2004). Characteristics and use acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8 (1), 31-54.
- Santiago, R. (1970). A contrastive analysis of some rhetorical aspects in the writing TESOL QUARTERLY 674 in Spanish and English of Spanish-speaking college students in Puerto Rico. *Dissertation Abstracts International*, 31(12), 6368A.
- Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic, and remedial capabilities in children's composing. In M. Martlew (Ed.), *The psychology of written language: A developmental approach* (pp. 67-95). London: Wiley.
- Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 778-803). New York: Macmillan.
- Schraver, K. (1992). Teaching Writers to Anticipate Readers' Need. *Written Communication*, 9 (2), 179-208.
- Silva, T. (1990). A comparative study of the composing of selected ESL and native English speaking freshman writers. *Dissertation Abstracts International*, 51(10), 3397A.
- Smith, L.M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. In L. Shulman (ed). *Review of Research in Education*. Itasca, III.: Peacock.
- Smith, N. (2003). Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In: Facchinetti, R., Krug, M., Palmer, F. (Eds.), *Modality in Contemporary English*. Mouton de Gruyter, Berlin/New York, pp. 241-267.
- Sommers, N. (1980). Revision Strategies of Student Writers and Experienced Adult Writers. *College Composition and Communication*, 31 (4), 378-387.
- Sommers, N. (1982). Responding to Student Writing. *College Composition and Communication*, 33 (2), 148-156.
- Sperling, M., & Freedman, S. (1987). A good girl writes like a good girl. *Written Communication*, 4, 343-369.
- Summer Smith, (1997). The Genre of the End Comment: Conventions in Teacher Responses to Student Writing. *College Composition and Communication*, 48 (2), 249-268.

- Stalker, J. W., & Stalker, J. C. (1988). A comparison of pragmatic accommodation of non-native and native speakers in written English. *World Englishes*, 7(2), 119-128.
- Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1, 217-233.
- Straub, R. (2000). The Student, the Text, and the Classroom Context: A Case Study of Teacher Response. *Assessing Writing*, 7 (1), 23-55.
- Tabachnick, B. G., & L. S. Fidell. (2001). *Using multivariate statistics*. Fourth edition. Boston: Allyn and Bacon, 2001.
- Takahashi, T., & Beebe, L. (1993). Cross-linguistic influence in the speech act of correction. In Kasper, G. and Blum-Kulka, S (eds.) *Interlanguage pragmatics*. Oxford: Oxford University Press.
- Treglia, M. (2009). Teacher-Written Commentary in College Writing Composition: How Does It Impact Student Revisions? *Composition Studies*, 37 (1), 67- 86.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147–170.
- Villamil, O., & De Guerrero, M. (1996). Peer revision in the second language classroom: Social cognitive activities, mediating strategies and aspects of social behavior. *Journal of Second Language Writing*, 3(1), 51–75.
- Villamil, O., & De Guerrero, M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, 19 (4), 491-514.
- Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.
- Wallace, D., Hayes, J. (1991). Redefining Revision for Freshmen. *Research in Teaching of English*, 25(1), 54-66.
- White, E.M., Lutz, W.D., & Kamusikiri, S. (1996). *Assessment of Writing: Politics, policies, and practices*. New York: Modern Language Association of America.
- Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International Encyclopedia of Education* (2nd ed., pp. 5738–5745). Oxford, UK: Pergamon.
- Wiener, H. (1986). Collaborative Learning in the classroom: A guide to evaluation. *College English*, 48, 52-61.
- Xu, G. (1990). An ex post facto study of differences in the structure of the standard expository paragraphs between written compositions by native and nonnative speakers of English at the college level. *Dissertation Abstracts International*, 51(6), 1942A.

- Yu, V., & Atkinson, P. (1988). An investigation of the language difficulties experienced by Hong Kong secondary school students in English medium schools: Pt. 1. *The problems. Journal of Multilingual and Multicultural Development*, 9(3), 267-284.
- Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly*, 16, 195-209.
- Zamel, V. (1985). Responding to Student Writing. *TESOL Quarterly*, 19 (1), 79-102.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4, 209-222.
- Zimmermann, R. (2000). L2 writing: subprocesses, a model of formulating and empirical findings. *Learning and Instruction*, 10, 73-99.

Vita

Hoonmil Kim was born and raised mostly in Seoul, Korea. At age nine, she moved to the U.S. with her family and lived there for three years, where she was first exposed to a second language and culture. She entered Yonsei University in Seoul, Korea, in 1990 where she received the degree of Bachelor of Art in Social Work in 1994. In 1996, she entered Graduate School of Interpretation and Translation at Hanguk University of Foreign Studies in Seoul, Korea to pursue a career as a language expert and was trained as a conference interpreter and translator for the languages of English and Korean. From 1998 to 2008, she worked as a successful conference interpreter and translator in various business fields, during which she worked as a lecturer at the Graduate School of Interpretation and Translation at Hanguk University of Foreign Studies. In September 2008, she entered the University of Texas at Austin and was in the program from 2008 until 20012. While at UT, she served a social chair for the student organization, Foreign Language Education Student Association (FLESA) from 2009-2010. She received Education Annual Fund Endowed Presidential Scholarship in Education for the 2010-2011 academic year and an academic scholarship from UT ESL Services for the 2011-2012 academic year. Her research interest includes writing development, vocabulary acquisition, and bilingual education. She lives in Seoul, Korea with her two kids and husband

Permanent address: Unit # 101-1802, Hyundai Apt. 53, Jamwon-dong, Seocho-gu, Seoul, Korea.

This dissertation was typed by Hoonmil Kim.

