THE EFFECTS OF WEB-BASED PEER REVIEW ON STUDENT WRITING

A dissertation submitted to the
College of Education, Health and Human Services
in partial fulfillment of the requirements
of the degree of Doctor of Philosophy
Kent State University

by
Ryan S. Wooley
December 2007
Copyright © 2007 by Ryan S. Wooley

All Rights Reserved
A dissertation written by

Ryan S. Wooley

B.A., University of Akron, 1995
M.A.T., Kent State University, 1997
M.Ed., Kent State University, 2004
Ph.D., Kent State University, 2007

Approved by

_______________________________, Director, Doctoral Dissertation Committee
David W. Dalton

_______________________________, Co-director, Doctoral Dissertation Committee
Christopher A. Was

_______________________________, Members, Doctoral Dissertation Committee
Shawn M. Fitzgerald

_______________________________,
Christian D. Schunn

Accepted by

_______________________________, Chairperson, Department of Educational Foundations and Special Services
Awilda C. Hamilton

_______________________________, Dean, College of Education, Health and Human Services
Donald L. Bubenzer
Some theorists consider writing and cognition to be symbiotic if not synonymous. Writing is regarded not only a means of communicating and assessing understanding of content knowledge, but as a way of constructing knowledge. Yet, though writing is likely a beneficial activity in most, if not all, disciplines, it has been difficult to implement in content courses. Robust online peer review systems for student writing now offer solutions to many of the problems that have impeded peer review activities in the past.

Research in self-explanations and reciprocal teaching has suggested that students stand to benefit cognitively by articulating explanations to self and others, but this research has been conducted primarily in math and science domains. There have been few, if any, investigations of the effects of articulating feedback for others on one’s own subsequent writing.

The goal of this study was to examine the effects of reviewing on one’s subsequent writing. Further, the study sought to illuminate distinctions between different types of reviewing and reviewer preparation, namely the effects of feedback elaboration and the effects of providing prototypical examples of helpful and unhelpful feedback.

Results indicate that students who provided elaborate forms of feedback, which included free-form comments, performed significantly better on their own writing than students who provided numerical ratings only. In this context, the use of examples did
not have significant effects on reviewers’ subsequent writing quality. Also, review-first groups did not perform significantly better than write-first groups, however, the author notes that the design of the study may have inadvertently counteracted reviewing effects.
ACKNOWLEDGEMENTS

I would like to thank several individuals who made substantial contributions toward the completion of this research project. First, I would like to extend my sincere appreciation to my advisor, Dave Dalton, who has been a remarkable mentor to me academically and professionally, and for his patient guidance through this dissertation process. I would also like to thank my other committee members—Dr. Chris Was, Dr. Shawn Fitzgerald, and Dr. Christian Schunn—for their generous investments of time, energy and expertise. Dr. Schunn deserves additional thanks for designing the SWoRD system, the sine qua non of this dissertation, and for reaching across institutional and geographic boundaries to participate in this project.

Second, I would like to thank Dr. Kwangsu Cho, who helped to create and administer the SWoRD system. During this project, he graciously fielded the increased demands created by treatment conditions and other idiosyncrasies of the study design. I would also like to thank each of the instructors who agreed to participate in the project: Dr. Anne Morrison, Dr. Larry Froehlich, Mary Louise Allen-Huffman, and Tom Foster. These instructors were willing to take the chance of engaging in a novel form of pedagogy, knowing there would be resistance from their students.

Finally, I would like to thank my wife Christina for her immense patience and support through this process.
TABLE OF CONTENTS

Page

ABSTRACT ........................................................................................................................... iv

ACKNOWLEDGEMENTS ................................................................................................... vi

LIST OF TABLES .............................................................................................................. x

CHAPTER

I INTRODUCTION ........................................................................................................... 1

Background ....................................................................................................................... 1
Rationale ........................................................................................................................... 8
Purpose of Study .............................................................................................................. 9

Research Questions and Related Hypotheses

Phase I .............................................................................................................................. 10
Phase II ............................................................................................................................. 10
Phase I .............................................................................................................................. 10
Phase II ............................................................................................................................. 10

Limitations of Study ...................................................................................................... 11
Delimitations of Study ................................................................................................... 11
Operational Definitions ................................................................................................. 12

II REVIEW OF THE LITERATURE .............................................................................. 14

Writing Curricula and Research ................................................................................... 15
Articulation .................................................................................................................... 24
Teach to Learn .............................................................................................................. 35
Measuring Writing Quality ........................................................................................... 42
Peer Review ................................................................................................................... 50
Feedback ....................................................................................................................... 56
Models and Modeling ................................................................................................. 62
Conclusion ..................................................................................................................... 67
III METHODS ........................................................................................................ 71
  Phase I ........................................................................................................ 71
  Phase II ........................................................................................................ 71
  Participants ................................................................................................. 71
  Apparatus .................................................................................................... 72
  Instrumentation ........................................................................................... 73
  Procedures .................................................................................................... 78
    Treatments .................................................................................................... 79
      Write First ................................................................................................... 80
      Review First, With Examples, With Elaboration ................................... 81
      Review First, With Examples, Without Elaboration .............................. 81
      Review First, Without Examples, With Elaboration ............................... 81
      Review First, Without Examples, Without Elaboration ......................... 82
  Experimental Design and Data Analysis .................................................. 82
    Research Hypotheses ................................................................................ 82
      Phase I ...................................................................................................... 82
      Phase II .................................................................................................... 82
  Design ........................................................................................................... 82
  Analyses ........................................................................................................ 84

IV RESULTS ..................................................................................................... 87
  Overview ...................................................................................................... 87
    Phase I—One Factor ANOVA .................................................................... 87
    Phase II—Factorial ANOVA ....................................................................... 88
      Preliminary Analyses ............................................................................. 88
      Hypothesis 1 ............................................................................................ 89
      Hypothesis 2 ............................................................................................ 90
      Hypothesis 3 ............................................................................................ 90
  Summary ....................................................................................................... 92

V DISCUSSION ................................................................................................. 93
  The Role of Context .................................................................................... 94
  Expected Results ......................................................................................... 100
  Unexpected Results ...................................................................................... 101
    Review-First Versus Write-First .............................................................. 101
    Examples .................................................................................................. 102
  Implications of Study .................................................................................. 107
  Limitations of Study ................................................................................... 108
    Draft One as a Measure ............................................................................ 108
    Time .......................................................................................................... 109
Maturity of Participants .........................................................109
Diffusion of Treatments ..........................................................110
Absence of Attitudinal Data ......................................................111
Recommendations for Future Research .................................111

APPENDICES ......................................................................................113
A. SWoRD Review Form ...............................................................114
B. Assignment Description ..........................................................119
C. Topic Choices .............................................................................123
D. Group 2 Instructions ................................................................125
E. Group 3 Instructions .................................................................129
F. Group 4 Instructions .................................................................133
G. Group 5 Instructions .................................................................137
H. Group 6 Instructions .................................................................141
I. Group 1 Instructions .................................................................145
J. Consent Form ...............................................................................149
K. Schedule Overview .................................................................151
L. Logistical Design .......................................................................153
M. Assignment Sheet – Examples Treatment ............................155
N. Adjusted Group Results ...........................................................157
O. SWoRD Score Results ............................................................162
P. MANOVA Results .....................................................................167

REFERENCES ....................................................................................180
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Correlations Between Expert Scores and SWoRD Scores by Dimension</td>
<td>76</td>
</tr>
<tr>
<td>2. IntraClass Correlation Coefficients (ICC) for Expert Scores</td>
<td>78</td>
</tr>
<tr>
<td>3. Means and Standard Deviations of Writing Scores by Treatment</td>
<td>88</td>
</tr>
<tr>
<td>4. Means and Standard Deviations by Factorial Cell</td>
<td>91</td>
</tr>
<tr>
<td>5. Factorial ANOVA – Main Effects and Interactions for Examples and</td>
<td>91</td>
</tr>
<tr>
<td>Elaboration</td>
<td></td>
</tr>
<tr>
<td>6. Content of Feedback Examples and Problems Noted for Them</td>
<td>104</td>
</tr>
<tr>
<td>N1. Adjusted Score Means and Standard Deviations by Factorial Cell</td>
<td>160</td>
</tr>
<tr>
<td>N2. Adjusted Score Factorial ANOVA – Main Effects and Interactions</td>
<td>161</td>
</tr>
<tr>
<td>O1. SWoRD Score Means and Standard Deviations by Factorial Cell</td>
<td>165</td>
</tr>
<tr>
<td>O2. SWoRD Score Factorial ANOVA – Main Effects and Interactions</td>
<td>166</td>
</tr>
<tr>
<td>P1. Means and Standard Deviations of Writing Scores by Treatment and</td>
<td>169</td>
</tr>
<tr>
<td>Writing Dimension</td>
<td></td>
</tr>
<tr>
<td>P2. Pearson Correlation Coefficients for Dependent Variables – Phase</td>
<td>170</td>
</tr>
<tr>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P4. Means and Standard Deviations for Logic Dimension</td>
<td>174</td>
</tr>
<tr>
<td>P5. Means and Standard Deviations for Insight Dimension</td>
<td>175</td>
</tr>
<tr>
<td>P6. Factorial MANOVA – Main Effects and Interactions for Examples and</td>
<td>176</td>
</tr>
<tr>
<td>Elaboration</td>
<td></td>
</tr>
<tr>
<td>P7. Factorial MANOVA – Main Effects for Examples and Elaboration</td>
<td>178</td>
</tr>
<tr>
<td>without Flow</td>
<td></td>
</tr>
</tbody>
</table>
P8. Factorial MANOVA – Main Effects for Examples and Elaboration without Logic ................................................................. 178

P9. Factorial MANOVA – Main Effects for Examples and Elaboration without Insight ................................................................. 179
CHAPTER I

Background

The practice of writing has long been linked to cognition. Writing is regarded by many as not only a means of communicating and assessing understanding of content knowledge, but a sophisticated process of constructing knowledge. In Zinsser’s (1988) words, writing is a “form of thinking” (p. vii). Various movements in education, most notably Writing across the Curriculum (WAC) and Writing in the Disciplines (WID) (Bazerman et al., 2005), have emerged from the premise that writing leads to generalized and specific knowledge. One of the most promising means of incorporating writing across the disciplines has been the use of peer review. Prior to the advent of robust web-applications seen in recent years, peer review in its traditional face-to-face form has been plagued with limitations: personality conflicts, favoritism, inefficient grading schemes, and consumption of class time.

Powerful online peer review systems, such as the SWoRD system (Scaffolded Writing and Re-Writing in the Discipline) developed at the University of Pittsburgh, have emerged with solutions to managerial issues, inefficiencies, and personality conflicts that have impeded peer review in the past. These developments, combined with fresh research in cognitive science and educational technology have begun to unlock the powerful potential of writing to learn in an online world. However, even traditional peer review has not been sufficiently studied (Haswell, 2005), let alone its newer technology-mediated forms. There is much to learn about how to employ peer review in this emerging context. This study was designed to determine whether undergraduate writing
improves as a result of providing feedback on peer papers, and further, what types of feedback conditions are most beneficial. In particular, this investigation sought to identify the effects of varying degrees of feedback elaboration and the effects of presenting prototypical examples of helpful and unhelpful feedback to peer reviewers prior to the review process.

For the past 30 years research on writing has increasingly made the case for curricular attention to writing across all disciplines. James Britton’s (1970) publication of *Language and Learning* drew much-needed attention to the probable connection between writing and thinking. His work launched the early phases of the writing across the curriculum movement that casts writing as an essential method of learning in every discipline. Britton suggested we “shape thought at the point of utterance” (1980). He argued that we construct knowledge through the *process* of writing and through the act of committing to articulations that become recursively refined through revision. Implied in this assertion, and stated explicitly by Kelly (1963), is the notion that commitment to a verbalization is the beginning step in a process of mental organization. Thus, early drafts of writing bear similarity to hypotheses of scientific inquiry. In each case, an assertion is made, tested, and revised if necessary in a process of meaning-making.

A decade after the publication of *Language and Learning*, as cognitive science was shifting from a focus on well-defined to ill-defined problems (Bereiter & Scardamalia, 1987), research in writing was beginning to support Britton’s theoretical foundation. As writing pedagogy moved from an emphasis on product to an emphasis on process, qualitative research methods like think-aloud protocol analysis, which involves
subjects explaining their processes as they work, emerged as the dominant mode of writing investigation. Through analyses of this type, researchers observed enough similarity among writers to construct reliable writing models such as those proposed by Flower and Hayes (1980) and Bereiter and Scardamalia (1987). Bereiter and Scardamalia (1987) conducted nearly 200 studies and concluded that there were differences between the writing processes of novice writers, who tend to engage in *knowledge-telling* activities and those of expert writers, who tend to engage in *knowledge-transforming* activities. These models, discussed in greater detail in Chapter II, ignited a writing process movement that guided writing instruction practices for two decades and more firmly established composition as its own field.

Yet, despite increasing support for the theoretical connection between writing and cognition, writing has not been given enough attention across the curriculum. One recent study reported that only one percent of students in the United States have effective writing skills, and 85 percent have only basic writing skills (National Assessment of Educational Progress [NAEP], 1998). Other studies have corroborated these findings, showing widespread neglect of writing outside of composition classes, especially in colleges and universities (Intersegmental Committee of the Academic Senates of the California Community Colleges, the California State University, and the University of California [ICAS], 2002; National Commission on Writing in American Schools and Colleges [NCWASC], 2003). Considering Bereiter and Scardamalia’s assertions that only advanced writers engage in knowledge transformation, all content areas seem to have a stake in student writing abilities.
Neglect of writing in content courses can be at least partially attributed to difficulties in implementing writing activities in the classroom as well as to prohibitively time-consuming grading methods. Traditional methods of writing instruction are taxing on instructors in the disciplines, who are primarily focused on their discipline-based concepts and skills. Technology has recently provided some inroads to this problem. One of the most promising means of incorporating writing across the disciplines has been the use of peer review, especially with the efficiencies and solutions to bias that have accompanied the development of online systems.

These technological and theoretical developments invite empirical testing of peer review systems. Though such systems may lower the hurdles associated with grading writing, little is known about whether the processes of peer review hold cognitive benefits for participants. In particular, does writing improve as students are asked to articulate feedback on peer papers? Intuitively, it seems likely that asking students to commit to evaluations of peer writing will reinforce their understandings of strengths and weaknesses in their own writing. Research findings outside the context of writing support these notions. For example, Chi and her colleagues at the University of Pittsburgh have identified cognitive benefits associated with self-explanations of students as they solve physics problems (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Several subsequent studies of self-explanation have suggested similar correlations with learning (Aleven & Koedinger, 2002; Bielaczyc, Pirolli & Brown, 1995; Chi & VanLehn, 1991; Chi, deLeeuw, Chiu & LaVancher, 1994; Ferguson-Hessler & de Jong, 1990; Neuman &
Schwarz, 1998; Pirolli & Bielaczyc, 1989; Pirolli & Recker, 1994; Renkl, 1997; Schworm & Renkl, 2006).

Self-explanations can be meta-cognitive, involving the analysis of one’s own thought processes, monitoring, and control strategies (Hacker, 1998). Articulating feedback on peer writing is similar because it requires the provider of feedback to analyze and explain the relative success of peer writing methods. For example, an evaluator might argue that the main point is not clear or that there is not enough support behind it. While analyzing the written product is not the same as analyzing the process of producing the written product, a completed writing project, by virtue of its length, is generally a rich record of its generative processes. As a product, writing leaves traces of its generative processes behind. Instructors, who are usually in the analytic role of evaluator, frequently deduce information about the student cognitive processes leading to the completed piece of writing. In other words, there is more to evaluating writing than a mere determination of correct or incorrect. If a reviewer is forced to identify a quality of writing, such as the importance of having a clear thesis, it is likely that the same reviewer will be careful to clearly identify theses in his or her own writing. In peer review, the reviewer assumes an analytic stance akin to that of the teacher and stands to gain insight as teachers do.

The cognitive benefits of peer teaching and cross-age tutoring have been well-supported in research since Cloward (1967) found that tenth- and eleventh-grade students with difficulty in reading and text comprehension benefitted from tutoring struggling fourth and fifth graders. Several subsequent studies produced similar results across a variety of contexts (Barbetta, Miller, Peters, Heron, & Cochran, 1991; Bargh &
Schul, 1980; Giesecke, Cartledge, & Gardner, 1993; Palincsar & Brown, 1984; Palincsar & Brown, 1986; Pigott, Fantuzzo, & Clement, 1986; Rekrut, 1992; Wheldall & Colmar, 1990; Wheldall & Mettem, 1985). In fact, peer teaching and cross-age tutoring have even been supported in meta-analyses (Britz, Dixon, & McLaughlin, 1989; Byrd, 1990; Cohen & Kulik, 1981; Cohen, Kulik, & Kulik, 1982). Leelawong et al. (2002) assert that teachers’ knowledge structures are organized as much by providing feedback as by preparing materials to teach.

Given the research findings on self-explaining and on peer teaching, it seems likely that asking students to articulate feedback on the quality of peer writing will lead to stronger abilities to discern qualities in one’s own writing, and by extension, stronger writing of one’s own. However, writing processes and sub-processes are complex and idiosyncratic. Results from other domains, such as the math and science problem-solving most often studied in self-explanation research, may have little application in the study of writing processes. Few, if any, researchers have examined the impact of articulating feedback for peers on one’s subsequent writing. Further, little is known about the effects of treatment variations in the context of preparing students to articulate feedback. For example, does providing more elaborate feedback lead to improved writing for the provider of feedback? Does the provision of examples of feedback help prepare the giver of feedback to reap the cognitive benefits of articulating feedback?

Feedback research outside the domain of writing has explored questions regarding elaboration and in general is favorable toward increased elaborations (Hancock, Thurman, & Hubbard, 1995; Kulhavy & Stock, 1989; Mason & Bruning, 2001;
Pridemore & Klein, 1995). Because these studies cross several domains, it seems likely that encouraging more elaborate feedback during peer review might lead to improvements in student writing. However, elaboration effects have not been examined in the context of peer review of writing. Just as with self-explanations, any assumptions that results transfer to the domain of writing are untested. Kulhavy and Stock (1989), whose feedback model is considered to be the most definitive early model, admit that the model is likely limited to simpler recognition task demands. Because writing is a more complex subject of evaluation, it is plausible that students who are asked to articulate feedback on peer writing might suffer the following: (a) exposure to examples of bad writing which negatively influences reviewers’ subsequent writing; (b) surpassing of one’s threshold of mental effort (Cennamo, 1989; Weiner, 1979). Finally, and perhaps most importantly, most feedback elaboration research focuses on the effects of elaboration on the receiver of feedback rather than on the giver of feedback. Clearly there are many reasons to more closely examine the effects of providing feedback to peers on one’s subsequent writing, especially with regard to feedback elaboration.

Another factor that might impact reviewing activities is the use of examples of effective or ineffective feedback. Examples research, not surprisingly, is generally supportive of the use of examples and modeling across several instructional contexts (Atkinson, Derry, Renkl, & Wortham, 2000; Charney & Carlson, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Cooper & Sweller, 1987; Nitsch, 1977; Pirolli, 1991; Sweller & Cooper, 1985). The provision of examples of effective and ineffective feedback may lead to improved cognitive organization during reviewing activities that
might translate to a student’s subsequent writing. However, only Charney and Carlson (1995) examined the effects of models in the context of writing, in particular investigating how students make use of model texts as they learn to write a method section for a simple psychology experiment. While no effects were observed for including varied quality models or their labels, significant differences in paper organization between the models group and the no-models group on organization were observed. While much remains to be learned about the effects of models on student writing, it appears that researchers might consider some of the dependent qualities of writing separately, perhaps in addition to assessing them holistically.

Rationale

Web-based peer review has not been available long enough to establish a deep research base. Research on traditional face-to-face peer review is sparse and has been conducted primarily in composition courses, revealing little about its use in content courses. Robust web-based peer review systems have only recently come into their own, opening the door for inquiry into their use across the curriculum to facilitate the writing process. Further, existing peer review research has tended to focus on the effects of reviewing on the receiver of feedback rather than effects of reviewing on the person who provides the feedback. Building on the theories of Britton (1970) and others whose work has suggested the value of articulation of ideas on one’s mental development and organization of conceptual ideas, more empirical studies are needed to make clear the connections between articulation of feedback and the development of one’s abilities to express ideas in writing. In addition, there is a need to explore distinctions between the
various types of feedback and the various methods for training students to give feedback to their peers. Based on related research in other domains, it appears likely that providing examples of prototypical feedback to reviewers might strengthen the reviewing experience and therefore subsequent writing, however use of exemplar feedback has not been studied in this context. Also, given results of feedback elaboration treatments in content fields and the self-explanation literature supporting the use of explanations, as well as research on the cognitive benefits of teaching, it seems likely that students will benefit cognitively from providing elaborate feedback to their peers, yet it is possible that such tasks might increase cognitive demands enough to elicit a mental effort threshold. In other words, there is a possibility that students will suffer from cognitive interferences or fatigue. There have been no empirical studies in the context of peer review of writing.

Purpose of Study

The purpose of this study was to examine differences in the quality of student writing between students who are asked to articulate feedback on peer papers prior to writing their own papers versus students who write without having provided feedback on peer papers. In addition, the study examined effects of two different treatments of review processes: (a) the inclusion of samples of helpful and unhelpful reviews in preparing reviewers for the review process; (b) the inclusion of comment articulation in addition to numerical rating of peer papers.
Research Questions and Related Hypotheses

Phase I

Does the articulation of feedback by reviewers in a web-based peer review system for writing result in a higher quality of reviewers’ subsequent writing?

Phase II

Does the provision of examples of helpful and unhelpful reviews result in a higher quality of reviewers’ subsequent writing? Does more elaborate feedback result in a higher quality of reviewers’ subsequent writing? Is there an interaction between examples and elaboration treatments?

This study will measure the following hypotheses:

Phase I

1. Articulation of feedback by reviewers using a web-based peer review system for writing will result in a higher quality of reviewers’ subsequent writing in an Educational Psychology class at a large, Midwestern state university.

Phase II

1. Exposing reviewers to examples of prototypical feedback will result in a higher quality of reviewers’ subsequent writing.
2. Providing elaborate comments in addition to simple numeric ratings will result in a higher quality of reviewers’ subsequent writing.
3. There will be an interaction between type of feedback and type of exposure to prototypical examples.
Limitations of Study

The study has a few key limitations. Due to the absence of pretest, there was no way to establish between-group equivalency or to measure the dependent variables absent all treatment conditions. However, since subjects were randomly assigned to groups, it is likely that individual differences were spread evenly across groups. The lack of a pre-assignment survey also prevented the researcher from gathering potentially significant historical information about subjects, such as previous experience with peer review processes, and particularly with the SWoRD system. Internal validity was also threatened by the lack of control for a diffusion of treatments effect. Specifically, there was no way of knowing if subjects shared information between groups that may have affected the results of the study. Finally, because the sample being studied was divided between four instructors, there was potential for a teacher effect, although this limitation was mediated by standardized, prepared packets of instructions for the assignment.

Delimitations of Study

Delimitations of the investigation were fairly straightforward. Since random selection from all education majors at all undergraduate colleges was prohibited by design, it is difficult to generalize any results beyond this one institution. Further research will be required to determine if similar results may be achieved in other colleges, other universities, other regions of the country, perhaps even to students of other educational levels.
Operational Definitions

1. *Asynchronous System*: a computer application that facilitates the activities of multiple participants at disparate moments in time.

2. *Consistency*: the extent to which any given reviewer systematically discerns good papers from poor papers. This is one of three dimensions that compose the review accuracy of reviewers in the SWoRD system (Cho & Schunn, 2007).

3. *Elaboration*: the inclusion of freeform feedback comments in addition to Likert ratings rendered by reviewers during review (see Appendix A).

4. *Examples*: instructor-provided models of prototypical student feedback comments of various quality levels (see http://peerfeedback.net).

5. *Experts*: external evaluators with expertise in composition and formal training in the evaluation of student writing.

6. *Flow*: the extent to which a paper is readable and comprehensible by readers (Cho & Schunn, 2007). This is one of the three dimensions of writing that SWoRD participants are asked to evaluate.

7. *Insight*: the extent to which a paper contributes new knowledge and insight beyond given texts and materials (Cho & Schunn, 2007). This is one of the three dimensions of writing SWoRD participants are asked to evaluate.

8. *Logic*: the extent to which a paper is logically coherent in terms of text structure that organizes various facts and arguments (Cho & Schunn, 2007). This is one of the three dimensions of writing SWoRD participants are asked to evaluate.
9. **Peer Review**: a method of evaluating student writing which places students in the role of reviewer and relies on the aggregation of student-generated writing scores.

10. **Peers**: classmates who are also participating in the SWoRD assignment.

11. **Spread**: the extent to which each reviewer distributes scores too narrowly or too widely across papers. This is one of three dimensions that compose the review accuracy of reviewers in the SWoRD system (Cho & Schunn, 2007).

12. **SWoRD**: “Scaffolded Writing and Rewriting in the Discipline,” an asynchronous online peer review system for writing in content courses.

13. **Systematic differences**: the extent to which any given reviewer is systematically generous or harsh in assessing papers. This is one of three dimensions that compose the review accuracy of reviewers in the SWoRD system (Cho & Schunn, 2007).

14. **Web-based**: the functionality of the system is dependent upon live Internet connectivity.
CHAPTER II
REVIEW OF LITERATURE

Literacy (reading and writing) has been a primary motivation for developing most educational institutions throughout history (Bazerman et al., 2005). Yet despite its centrality, educational institutions have long struggled to position writing within the curriculum. Should writing be taught in a composition classroom, where it can be studied and practiced independently? Or should writing skills be cultivated in every classroom since it is so closely tied to learning that it must have a place in all curricula? While current research increasingly makes a case for writing across the curriculum, and despite movements such as Writing Across the Curriculum (WAC) (Bazerman et. al, 2005) and Writing in the Discipline (WID)( Bazerman et. al, 2005), writing has been neglected outside the composition classroom (ICAS, 2002).

While this dissertation deals with a particular aspect of writing –the effects of writers articulating feedback on peer work—and even more specifically within an online peer review system, a sufficient review of the literature must first address broader questions, including: Why should writing be embedded across the curriculum? Why facilitate writing with peer review? Why use an online system over face to face? These questions provide a context for considering the specific questions of the current study. Therefore to adequately frame the study, this review will be multidimensional, cutting across several domains, each with its own substantial foundation of research. This investigation will consider only the relevant aspects of each vane of research, while building sufficient context for the current study. Relevant literature includes, but is not
limited to: (a) writing curricula and research, (b) articulation (explaining to self and others), (c) cognitive benefits of teaching, (d) peer review and evaluation of writing, (e) feedback, and (f) observational learning/modeling.

Writing Curricula and Research

The history of writing curriculum in American higher education at least partially explains why writing was late to adopt cognitive research methodology. Before the late 19th century, higher education was largely rhetorical and aimed at producing a “religious and secular elite” (p. 14, Bazerman et. al, 2005). Education was not necessarily related to the development of vocational skills. Instead, practical training for professions happened through apprenticeship during this period. Two events would revise this conception of the Academy: First, the Morrill Act of 1862, also known as the Land Grant College Act, was intended to help establish public institutions in every state that would be available to those in all social classes and that would be oriented to professions preparation, including agriculture, mechanical arts, and home economics. (Lightcap, 2007). Second, the opening of Johns Hopkins University in 1876, modeled after German research universities, introduced the concept of specialized departments aimed at generating domain-specific knowledge. Since, writing and rhetorical study were not recognized as disciplines in their own right, they were often not included as separate departments.

As accessibility to education was increasing in the late 19th century, neglect of literacy and writing would become immediately evident, leading to the first of many literacy crises in education (Bazerman et. al, 2005). Smagorinsky (1995) notes that whenever institutions of higher education have been opened to more diverse student
bodies, proclamations of literacy crises have followed, leading, inevitably to increased scrutiny of the handling of writing and rhetoric. The crisis of 1875-1885 saw the mandatory freshman year composition course as a response (Bazerman, et. al, 2005). The doubling of enrollments at American universities between 1920 and 1930 led to another literary crisis and subsequent questioning of the effectiveness of first-year writing efforts. As college enrollments in the United States again doubled in the decade of the 1960’s (US Census, 2007) challenges of dealing with literacy and writing in the curriculum again bubbled to the surface, particularly as increasing numbers of students were entering higher education with questionable preparedness. For example, after implementing a new open admissions policy, the City University of New York (CUNY) focused on the challenges of writing by developing innovative writing curricula (Bazerman et al., 2005; Fullinwider, 2002; Yood, 2005) after their open admissions policy, born in 1969, opened the doors to a larger and more diverse body of students (Fullinwider, 2002). Several public universities adopted similar admissions policies, faced similar challenges, and turned attention toward literacy skills and new writing curricula.

By the 1960’s, due to its continued independence in the curriculum and its recurrence as the focus of attention in higher education, composition was emerging as its own academic discipline. This elevated status manifested itself in the birth of new dedicated academic journals and conferences (Connors, 1995). The month-long Dartmouth Conference of 1966, which brought British, Canadian, and American educators together (Bazerman et al, 2005; Watson, 1992) is generally regarded as the critical event in establishing composition as a field. At the seminar, participants
discovered that the British and American education systems were on divergent trajectories, with American education shifting away from its progressive movements begun in the 1920’s and 1930’s and British education moving in the much the opposite direction, toward more progressive attitudes (Peritz, 1994). In essence, American participants were influenced by British participants and began importing writing theories and curricular developments. Ironically, given the newly established independent status of the field of composition, momentum created at the 1966 conference inspired a “language across the curriculum” movement that would ultimately become known as Writing Across the Curriculum (Peritz, 1994) and would be characterized by the integration of writing in all disciplines. James Britton, a British conference participant, became a key figure in the movement, launching some of the earliest research projects aimed at defining the connection between writing and learning. In 1970, he published his influential book, Language and Learning, he argued that writing was a form of thinking (1970). It would take a decade for these ideas to take hold and organize writing research and practice (Best, 1995).

Writing to Learn (Zinsser, 1988) is an ideology that has influenced the study of writing as much as its curricular residence; it is the conceptual bridge to a cognitive science orientation to studying writing that seeks to connect writing, thinking, and learning (Best, 1995; Hayes & Flower, 1986; Kintsch, 1987). Informed by this perspective, writing researchers in the early 1980’s began to focus on mental processes over written products in an attempt to organize the cognitive sub-processes involved in text creation. Think-aloud protocol analysis emerged as the predominant means of
analyzing the mental activities of writers. In this method, researchers would ask students to explain what they were thinking as they worked through a writing task. The technique renders large amounts of data that can then be analyzed for patterns. Such patterns formed the basis of Flower and Hayes’ (1980) landmark cognitive model of the writing process, regarded by many as the most significant early guidepost in the transition to a process orientation to writing. The model portrays writing as a complex recursive process of several concurrent sub-processes and constraints, a revision of earlier conceptions of writing as a linear series of discreet steps. The model suggests that writing involves significant cognitive activity, requiring frequent switching between the mental sub-processes of planning, text generation and editing. The Flower and Hayes (1980) model gave writing researchers and instructors a new language with which to discuss and design writing instruction. The collaboration of Flower, a university writing program director, and Hayes, a cognitive psychologist, was symbolic of a larger, emerging symbiosis.

Bereiter and Scardamalia (1987), a prolific team of cognitive psychology researchers who were also dedicated to studying the writing process in the early 1980’s complemented the work of Flower and Hayes by focusing more on control strategies and memory retrieval (Kintsch, 1987). Bereiter, Scardamalia, and their colleagues completed about 120 studies over the eight years leading to the publication of The Psychology of Written Composition (1987), in which they proposed their own writing models. They observed “an interesting bifurcation in the literature between treatments of writing as a difficult task, mastered only with great effort, and treatments of it as a natural consequence of language development” (p. 5, 1987). Believing that there was compelling
evidence for each point of view, they were open to both perspectives when analyzing their data and would ultimately propose two distinct writing process models: one based on the “psychology of the natural” (p.5) and the other based on a “psychology of the problematic” (p.5). Rather than providing competing explanations of the cognitive activities of the writing process, the two models accounted for differences between mature and immature competence in writing. Less sophisticated writers are more likely to follow the *knowledge-telling* model emerging from a psychology of the natural, where advanced writers tend to follow a *knowledge-transforming* model based on the psychology of the problematic. The authors note a similar contrast between everyday thinking and formal reasoning, noting the distinguishing feature of more studied abilities “is that they involve deliberate, strategic control over parts of the process that are unattended to in the more naturally developed ability” (p. 6). While the sophisticated mental control processes might increase cognitive load and therefore the potential for error, they also significantly increase one’s capacity for acquiring expertise (1987). In the knowledge-transforming model, advanced writers must frequently and recursively switch between the simultaneous challenges of a *content problem space*, involving what the writer knows or needs to know about a topic, and a *rhetorical problem space*, involving what a writer knows or needs to know about writing within a particular context.

The process models of Flower and Hayes (1980) and Bereiter and Scardamalia (1987) have informed writing instruction, including WAC efforts, for nearly thirty years, but are not without their limitations. Yood (2005) claims the central tenets about the writing process are fully known and should no longer be the dominant focus in writing
research, but much goes unexplained. For example, there have been relatively few investigations exploring individual differences among writers or the external variables and treatments having the greatest effect on learning (Best, 1995). Because of the emphasis on process and the need to describe prototypical activity, writing research has been appropriately dominated by qualitative methods (1995). However, limiting writing research to qualitative methodology has hampered the generalization of results and has often discouraged tighter, more precise controls associated with quantitative methods. In short, research in writing process is ripe for a transition from broad, replication-prohibitive qualitative methods to the more granular and reliable analysis of experimental research. Protocol analysis has revealed consistent patterns in student writing processes, but it does not provide the necessary controls for a deep analysis of relevant variables. For example, the role of feedback in the recursive processes of writing has been well-studied, but more precise research might reveal more about how different types of feedback function across various content domains and educational levels.

Of course, interpretations of the current state of writing research are as varied as proposed trajectories. Best (1995), for example, argues that qualitative methods, especially protocol analysis should continue as the dominant approach, but that it should be used to investigate other types of articulation beyond just writing text. She claims that process approaches to instruction are incongruent with the quantification of student performance demanded by most school environments, but does not provide a suggestion for how to bridge this gap. Conversely Hatch (1992) criticizes the protocol method as an experimental method that contaminates its own sample. He argues that the act of
verbalizing processes, especially when prompted by a researcher, is not representative of the true mental processes of writers. Hatch (1992) is especially critical of the metaphor that equates writing and thinking, noting that many intelligent and articulate students struggle with the act of writing. He seems to have confused process models as resting on fixed, innate learner qualities rather than specific, learnable, skill-based forms of thinking. Ransdell and Levy (1996) describe writing as “a process that requires extensive self-regulation and attentional control” (p. 93,). This description is more aligned to current views of intelligence as a system of malleable control strategies (Dweck, 1999). It seems Hatch’s (1992) call for a behavioral perspective to guide future writing research has been flatly ignored.

Other views of the direction of the study of writing have been better received by teachers and scholars. Yood (2005) is one of several compositionists to have declared an end to the process paradigm, citing “writing-as-public” (p. 5) as the new focus of composition study. In fact, Thomas Kent (1999) published a collection of articles entitled Post-Process Theory: Beyond the Writing-Process Paradigm, that suggests three assumptions of the post-process movement: that writing is (a) interpretive (involving sense-making), (b) situated (rooted in context), and (c) public (a social process involving multiple language users). A critical review of these tenets, however, reveals that they are not necessarily at odds with the ideology of process theories, which generally account for all three (Porter, 2000). What seems to organize post-process ideology is not so much a rejection of process, but recognition that its current course has reached its useful limitations. Post-process critiques tend to question empiricism as an appropriate
framework for understanding writing, based on the argument that there is no singular writing process, no predictability to writing, and no statements that can apply to all or even most writers or writing situations (Olson, 1999; Porter, 2000). Post-process theorists question the utility of what they consider to be overly general writing models that are too reductionist to offer any real insight (Porter, 2000). In other words, they suggest that analyzing and studying narrow aspects of writing in controlled contexts detaches those aspects of writing from complex, holistic influences of authentic writing activities that would otherwise be at work.

While the field may have outgrown the basic structures of Flower and Hayes’ (1980) and Bereiter and Scardamalia’s (1987) models, their potential in organizing composition research has not been fully realized. These models were intended as frameworks to guide contextualized research and pedagogy. Perhaps, then critiques of their limitations may be misdirected. As Porter (2000) points out, no one has ever learned to write by studying the models. Without models, writing is cast as entirely contextual, with no predictability or commonality cutting across contexts or domains, leaving little reason to study it or, for that matter, teach it. Conceived this way, writing abilities rest on mysterious, innate talents rather than on learnable control strategies. Considering that writing is essential to success in most fields, its learnability, in all of its various contexts, is foundational to the field of composition and to the enterprise of education itself.

The notion that writing skills can be learned is congruent with emerging evidence suggesting intelligence is changeable. Dweck (1999) calls this conception an “incremental” view of intelligence and contrasts it with the more traditional “entity” view
of intelligence. *Entity* views of intelligence conceive it as a stable, natural endowment, where *incremental* views of intelligence conceive it as resting on improvable control strategies. There is increasing support for incremental views of intelligence (Ceci & Liker, 1986; Dweck, Chiu, & Hong, 1995; Lave, 1989; Sternberg, 1998, 1999). If it is possible to extract generalizable strategies to promote intelligence, it is more than plausible that writing skills can be improved by teaching the control strategies that emerge through an empirical, social scientific approach to the study of writing. Such development depends upon understanding and making use of generalizable characteristics of writing processes. As compositionists are fond of noting, writing is contextual or *situated*, but anything that has a context, by definition, has a reference or relationship to commonality. There are generalizable aspects of writing that need to be studied and understood.

*Writing in the Disciplines*, a more contextual subset of WAC (Bazerman et al., 2005) is based on recognition of the importance of both the common and the contextual. According to Bazerman et al.:

When student writing is considered as the product of a particular disciplinary environment, expectations of student writing, and the subsequent response, assessment, and evaluation are more varied across disciplines, and there is a higher degree of consistency among teachers within a given discipline (p. 122, 2005).

Implicit in the statement is the idea that consistency across instructors is beneficial to students and that useful, generalized principles can emerge through appropriate research.
This notion further supports the need to move toward quantitative methodology to validate emerging cross-curricular approaches to writing. Writing is indeed “public,” and publicity is almost by definition normative.

Articulation

Of all of the strands of research relevant to the current investigation, the most foundational theoretically and empirically is the literature of articulation. This strand of research concerns the cognitive consequences of developing raw thought into language, of representing notions symbolically. The articulation literature has seen meaningful contributions from philosophers and scientists, creating a rich symbiosis of theory and practice. The current study investigates the effect of articulating feedback for peers on the quality of college students’ writing. Both theoretical and empirical literature suggest that articulation of analytical feedback on peer writing could lead to improved writing skills, although articulation has not been examined empirically in this particular context. The following paragraphs present a rationale for these assertions, beginning in the theoretical.

German philosopher Ernst Cassirer (1944) asserted that the most-defining characteristic of human beings is not their ability to reason, as ancient philosophers had suggested, but rather their use of symbolic thought and behavior. Cassirer suggested that all human language is metaphorical. The use of symbols and metaphors (language) to discern and refine meaning is distinctly human and so central to the human psyche that language and thought are considered inseparable (1944).

Others have noted the relation between linguistic expression and cognition. Zinsser calls writing a “form of thinking” (1988, p. vii) and equates clear writing to the
“logical arrangement of thought” (1988, p. viii). Zinsser describes the connection between writing and thinking in personal terms:

I thought of how often as a writer I had made clear to myself some subject I had previously known nothing about by just putting one sentence after another—by reasoning my way in sequential steps to its meaning. I thought of how often the act of writing even the simplest document—a letter, for instance—had clarified my half-formed ideas. Writing and thinking and learning were the same process. (p. ix)

Zinsser’s description focuses on process in an attempt to discern precisely how writing and thinking are connected. In Zinsser’s view, linearity and sequence are important elements in the writing process.

Although presented separately here and having a literature independent of learn by teaching or peer teaching, learner articulation should be recognized as an activity embedded in peer teaching. Teachers, after all, explain concepts, processes, and phenomena. The theoretical foundation of learner articulation, at least as it resonates in the field of education, hearkens back to James Britton’s important work, Language and Learning, published in 1970. Building on the work of Vygotsky (1986), who suggested that human development depended on sign systems—the symbols that help people within a culture to think, communicate and solve problems and drawing on his experiences as a researcher, teacher, and parent, Britton (1970) theorized that language is the chief means of organizing the world, and therefore a critical cognitive asset. Britton proposed that the articulation of existing knowledge is an essential priming mechanism for acquiring new
knowledge. Thinking, and therefore understanding, is shaped, sharpened and
strengthened through the act of utterance, both before and after exposure to new
information. In fact the Socratic method of teaching rests on this principle (Ploetzner,
Dillenbourg, Praier, & Traum, 1999).

Britton emphasized the recursive nature of articulation, a notion that had been
previously suggested by psychologist George Kelly (1963). In Kelly’s view, articulation
was akin to the scientific method, where one formulates hypotheses, predicts, tests, and
then revises hypotheses:

Man looks at his world through transparent patterns or templates which he creates
and then attempts to fit over the realities of which the world is composed. The fit
is not always very good. Yet without such patterns the world appears to be such
an undifferentiated homogeneity that man is unable to make any sense out of it.

Even a poor fit is more helpful to him than nothing at all. (pp.8-9).

Kelly’s description proposes two important aspects of the articulation-cognition
connection: (a) that articulation puts the speaker into a recursive process of meaning-
making, and (b) that the process involves the use of patterns. In a review of several
historical and empirical studies, Margolis (1987) suggests all thinking and judgment can
be reduced to pattern recognition.

Assuming a reasonable degree of soundness in theories that connect articulation
with cognition, educators in all domains might consider orchestrating learning events that
include articulation tasks. Some researchers have noted that “learner articulation,”
Koschmann & Lebaron, 2002) might bear similarity to anatomists’ conception of the
term *articulation* in the sense that several separate elements are fit together “to form an integrated whole.” (p. 250). In each case, articulation is an essential component in realizing the whole. The notion that the articulation of thought leads to better-organized and integrated thinking is supported empirically as much as it is intuitively compelling. Indeed an emerging body of research is supportive of this theoretical foundation. Often this research is not classified as articulation research per se, but rather under the categories of *self-explanations* or *explaining to others*. Further, most explanation studies have been performed in content areas such as science and math that are not often associated with high levels of linguistic articulation.

A recent review of the literature of explaining reveals that self-explaining is more studied than interactive explanation and that there are not clear advantages of using one form over the other (Ploetzner, Dillenbourg, Preier, & Traum, 1999). Explanations can be categorized into five levels of interactivity: (a) explaining to oneself, (b) explaining to a passive and anonymous listener, (c) explaining to a passive listener, (d) explaining to a constrained responder, and (e) mutual explanation (1999). These distinctions are based on the embedded characteristics within each type of explanation, such as audience awareness, and the ability to structure and steer interactions.

Several studies of self-explanation have suggested a strong correlation with learning—either through creation of new knowledge or through the refinement of existing knowledge—across the following levels and domains: (a) college physics (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi & VanLehn, 1991; Ferguson-Hessler & de Jong, 1990), (b) college computer programming (Bielaczyce, Pirolli & Brown, 1995;
Pirolli & Beilaczyc, 1989; Pirolli & Recker, 1994), (c) middle school health (Chi, deLeeuw, Chiu & LaVancher, 1994), (d) college probability (Renkl, 1997), (e) college analytical reasoning (Neuman & Schwarz, 1998), (f) high school geometry (Aleven & Koedinger, 2002), and (g) entry teacher science instruction (Schworm & Renkl, 2006).

Perhaps the significant early study on self explanations, Chi et al. 1989, illuminates the core issues of explanation inquiry. Chi et al. define learning as “a constructive process in which a student converts words and examples generated by a teacher or presented in a text, into usable skills, such as solving problems.” (p145, 1989). Since examples (discussed in detail in a separate section of this review) are often used to teach concepts, the study was designed to investigate ten college students taking a college physics course, where studying worked-out examples is common. The researchers, positing that students come to understand an example by explaining it, overcoming its incompleteness, and making inferences from it, asked students to study worked-out examples of physics problems and talk aloud as they reasoned through a self explanation.

This explanation activity was preceded by reading three relevant textbook chapters and a subsequent criterion test. Students had to demonstrate comprehension of prerequisite knowledge before proceeding to the new material. The new material included reading a chapter on particle dynamics, studying three pre-solved problems, and solving 25 problems. Participants were asked to talk aloud while studying example problems and then solving the 25 physics problems. These self-explanations were recorded and later analyzed. One student could not solve any problems and was eliminated from the pool. The rest of the participants were divided post hoc into two
groups: the four top performers and the four lowest performers. Data for the middle student was eliminated to keep group sizes equal. Because all of the students had performed equally well on a pre-test, high-scoring students, with mean success of 82%, 96%, and 68% depending on the type of problem, illustrated clear gain differences over low-scoring students, who had mean success rates of 46%, 62%, and 30% respectively. Verbalizations from these groups were classified via protocol analysis, revealing that: (a) high-scoring students gave significantly more self-explanations while studying worked-out examples, (b) strong performers uttered more accurate self-monitoring, metacognitive statements, and (c) strong performers did not refer back to examples as often as weak performers during problem-solving, suggesting they had better retention of the material.

While the Chi et al. study (1989) showed strong correlation between explanation behaviors and performance, its execution revealed some untested assumptions and some questions for further inquiry. For example, some researchers point out that the classification of self explanation might be misleading because participants were prompted to explain by the original researchers, who themselves could be considered an external audience. This observation emphasized the role of audience in explanation inquiry, leading subsequent researchers to attend to the various levels of interactivity as categorized by Ploetzner et al. (1999) above. Chi and colleagues (1989), with their post hoc analyses, sought to observe the natural activities of students, at least to the extent possible in an empirical study. This approach created a reference point for future researchers to examine various levels of instructional initiation of self-explanations. For
example, Chi, de Leeuw, Chiu, & LaVancher (1994) investigated the effects of systematically initiating self explanations and found significantly more gains in the problem-solving performance of groups of eighth grade students who were prompted to generate a large number of self explanations. The Chi et al. (1989) study also illustrated that analysis of what students articulate is at least as important as how often and how elaborate they articulate. Finally, the study prompted questions about the role of examples in informing student explanations.

After the original Chi et al. (1989) study showed a clear correlation between performance and self-explanation behaviors, several subsequent investigations refined the features of explanation literature. For example, Chi & Van Lehn (1991) focused on the content of self-explanations, helping them to draw a distinction between self explanations, which are generated in the context of learning something new, and elaborations, which involve the use of existing knowledge. Pirolli and Recker (1994) demonstrated that the longer students worked on an example, the fewer explanations were rendered, suggesting that self-explanations might have diminishing returns. Bielaczyc, Pirolli, and Brown (1995) found that people trained in specific explanation strategies showed greater regulation behaviors and consequently greater performance gains. Renkl (1997) addressed some of the confounds from previous research, such as time on task and quality of learning processes and found that there are qualitative differences in student-generated explanations, and that such differences (e.g. anticipative reasoning and principle-based explanations vs. passive and superficial explanations) account for variation in performance (1997).
More recently, some researchers have manipulated various ways of eliciting self explanations, from computer-based learning environments with embedded video (Renkl & Atkinson, 2002) and the use of a cognitive tutor, an intelligent instructional software program that supports guided learning by doing (Aleven & Koedinger, 2002) to representational tools such as argumentative maps and pro-con tables (Schwarz, Neuman, Gil & Ilya, 2003). In addition to expanding the collective repertoire of explanation-based instructional methods, researchers have also expanded the concept of articulation beyond traditional oral and written forms to include interactive physical gestures. Each of these strands of inquiry has contributed to the evolution of articulation literature, increasingly demonstrating the qualitative differences between instructional applications of explanation activities.

Explanation activities are meta-cognitive—that is, involving the understanding of one’s own thought processes, monitoring, and control strategies (Hacker, 1998). Hacker and Dunlosky (2003) argue there is a range of utility associated with meta-cognitive activities. With the models provided by Nisbett and Wilson (1977) and Ericsson and Simon (1980), Hacker and Dunlosky (2003) constructed a framework for categorizing verbalizations and their relative implementations in instructional activities. They summarize and describe three levels of verbalization and three types of verbal report. The three levels of verbal report are: (a) verbalization of verbal contents (e.g. “Keep talking about what you’re thinking as you read the story”), (b) verbalization of non-verbal contents (e.g. “Talk aloud as you work the Rubik’s Cube”), and (c) verbalization of verbal or non-verbal contents that involve explaining those contents (e.g. “Explain how you
make use of the rhetorical structures of the story as you read.”). Note that this third level requires an analysis of active control processes rather than merely a description of what one is doing. These three levels of verbalization occur across three different time orientations, what Hacker and Dunlosky (2003) call *types of verbal report*. The three types of verbal report are: (a) *concurrent* (reporting as one is thinking), (b) *retrospective* (reporting from past experiences), and (c) *prospective* (reporting current thoughts or remembered experiences for making predictions).

Hacker and Dunlosky (2003) argue that neither verbalizing verbal contents, what they call *level 1 verbalizations*, nor verbalizing non-verbal contents, what they call *level 2 verbalizations*, impact student thought processes, because these two types of verbalization lack the analytical component present only in *level 3 verbalizations*, where there is reflecting on information currently in verbal and nonverbal form *and* additional explanations and justifications for it. These more complex and analytical verbalizations are likely helpful in problem solving across all three time orientations, but research has primarily been focused only on concurrent verbalizations. This research has strongly supported the use of concurrent verbalization that requires analytical explanations.

Hacker and Dunlosky (2003) describe how professors might encourage level 3 concurrent verbalization:

Instructions that promote level 3 concurrent verbalization (for example, “Explain how you make use of the rhetorical structures of the story as you read”) force students to go beyond simply reporting what they are thinking. They must
deliberately change the course and structure of their thoughts as they verbalize responses to the instructions. (pp. 75-76)

With these techniques, students are asked to articulate their motives and the reasons behind their thoughts during problem solving. For their examples, the authors extend beyond science and math problems, expanding the usual focus of explanation research. They specifically describe a metacognitive activity aimed at the ill-structured task of writing a persuasive essay (2003). An ill-structured problem is one that can have many different successful solutions.

Of course, researchers familiar with Bereiter and Scardamalia’s (1987) knowledge-transforming model of advanced writers will recognize that writing itself is thought to have a dual problem space similar to that described by Hacker and Dunlosky as a Level 3 concurrent verbalization. The knowledge-transforming model draws separate cognitive processes for dealing with content and rhetorical control. In the same way that a writing task asks an advanced writer to simultaneously engage in both a content problem space (involving what students know, need to know, and need to employ about a given topic) and a rhetorical space (involving what students know, need to know, and need to put into action about how to write), Hacker and Dunlosky’s (2003) level 3 concurrent verbalization tasks ask students to engage in two separate but connected spheres: (a) conceptual content knowledge, and (b) knowledge of ones own motives and mental processes. Perhaps the analogy only works to the extent that students are aware of their engagement in the rhetorical problem space. Hacker and Dunlosky use the following writing prompt to elicit level 3 explanations: “Explain how you make use of the rhetorical
structures of the story as you read,” (pp. 75-76, 2003). Perhaps an equivalent prompt for
writing might be “Explain how you employ rhetorical devices to make your point.”

Yet, if writing is at least a partial record—a product—of the mental processes
involved in its creation and a learner’s engagement in the dual problem spaces of content
and rhetoric, it might provide a means for learners to make post hoc comparisons
between their own cognitive activities and those of their peers. Through peer review
activities, such comparisons would likely add strength and depth to one’s own
understanding of his or her cognitive processes during writing. The interactive dialogue
students would have about their writing, even if asynchronous and anonymous, could
potentially lead to more opportunities for learning than self explanation activities (Webb,
1989). For example, in any type of dialogue, a receiver of feedback is likely to
(a) identify missing information or inconsistencies, (b) push for clarification, and (c) pose
alternative points of view. This dialogue should, in turn, force the explainer to find new
information or restructure existing information (Ploetzner et al., 1999). Writing activities,
particularly when facilitated by a peer review process, give multiple opportunities for
various forms of articulation, including those consistent with Hacker and Dunlosky’s
characterization of the most effective verbalizations.

Almost universally, articulation research has shown positive effects from the
elicitation of explanations. However much of this research has been performed in
learning situations involving scientific or mathematical problem-solving and have
measured the effects of self-explanation on subsequent abilities to solve similar problems.
Few, if any, empirical studies have examined the use of explanations in verbal domains
such as writing. More specifically, there have been no investigations of this nature in the relatively new instructional domain of online peer review systems. On one hand the nearly universal affirmation of explanation prompting in the current literature suggests that its benefits might emerge in many instructional domains, yet recent refinements are showing that the context in which explanation activities occur impacts the qualitative differences among explanations—differences that ultimately transfer to performance. More information is needed about what specific combinations of explanation activities and instructional contexts are most effective, especially in the seldom-studied domain of student writing.

Teach to Learn

Although they have not been substantiated empirically, common estimates of retention suggest we learn: 5-20% of what we read or hear, 20-50% of what is demonstrated, 60% of what we write, 50-70% of what we discuss, 75-80% of what we experience or practice by doing, and 90-95% of what we teach (Betrus & Januszewski, 2002; Schunn, 2004; Uno, 1999). If these estimates are even remotely accurate, instructors of all disciplines may have reason to find ways of placing students in the teacher role.

The notion that students stand to gain cognitively from teaching their peers is not new. Kalkowski (1995) claims “peer and cross-age tutoring have been part of human existence since hunter-gatherer times” (p.1). Jenkins and Jenkins’ (1987) assert “Tutorial instruction (parents teaching their offspring how to make a fire and to hunt and adolescents instructing younger siblings about edible berries and roots) was probably the
first pedagogy of primitive societies” (p. 64). Bargh and Schul (1980) point out that the early one-room schoolhouses of New England often engaged older students in teaching the younger students, in part, to reinforce the older students’ understanding of material. Further, at least one reason graduate students have traditionally been required to teach is to facilitate mastery of material (McKeachie & Kulik, 1975).

Despite its long-standing position in American Education, teaching to learn is described by Weigel (2003) as “The Next Big Thing.” Weigel’s assertion underscores the momentum that has been generated by advances in computer technology. What some see as a difference in degree, Weigel sees as a difference in kind. Weigel points to the advantages of Groove and other networked and web-based software in establishing the necessary environments for the teach-to-learn model to thrive. Certainly an equally important factor is the maturity of literature related to the benefits of peer and cross-age tutoring.

Since the late 1960’s, educational research has supported claims of the cognitive benefits of peer teaching and cross-age tutoring. For example, Cloward (1967) found that a group of 240 tenth and eleventh-grade students with difficulty in reading and text comprehension showed significant gains in their ability to read and comprehend texts after tutoring struggling fourth and fifth graders for a period of about five months. Bargh & Schul (1980) found that undergraduates in an introductory psychology course who prepared to teach others to take a quiz on a passage showed better retention than those who prepared to take the quiz themselves. The results were attributed to the fact that students studied the material differently when they would be teaching others. Preparation
for teaching may have led to more organized cognitive structures for the teaching students.

Palincsar and Brown (1984) developed and tested an instructional method called reciprocal teaching, in which an adult tutor and an inexperienced student take turns discussing a text. The text is chosen by the tutor, who at first models teaching behaviors such as summarizing, questioning, clarifying, and predicting. Gradually, the inexperienced student is asked more frequently to assume the teacher role until the relationship functions more like a partnership. This method was taught to one of two experimental groups, the other receiving more traditional instruction. The study also included two control groups, one to control for the effects of repeated testing, and the other to participate in only pre and posttest taken by all participants. Twenty-four seventh-grade reading students were included in the investigation. After 20 days of instruction, students in the reciprocal teaching treatment showed significant improvement in text comprehension abilities over all of the other groups. Students in this group also showed improvements in their teacher behaviors, increasingly drawing out the main ideas of the text and posing questions and summaries in their own language. Most significantly, reciprocal teaching group performance on daily tests of text comprehension improved from an average of 30% correctness at the beginning of the study to an average of 80% correctness after only 11-12 days of instruction. (Palincsar & Brown, 1984).

Similar results have been found in other disciplines. Britz, Dixon, & McLaughlin (1989) reviewed peer tutoring and mathematics research from 1980-89. They concluded that peer tutoring usually lead to significant cognitive gains for low-achieving, mildly
handicapped, or socially disadvantaged children. Another mathematics study suggested that peer tutoring in an elementary math class led to significantly improved skills for the tutors (Pigott, Fantuzzo, & Clement, 1986). In the language arts, Rekrut (1992) compared story grammar posttests of high school students who either did or did not tutor fourth and fifth grade students on story grammars (diagrams of relationships between story elements) and found that the group that tutored performed significantly better. Several other studies have suggested achievement gains for tutees in the language arts. See, for example: Barbetta, Miller, Peters, Heron, & Cochran, 1991; Giesecke, Cartledge, & Gardner, 1993; Palincsar & Brown, 1986; Wheldall & Colmar, 1990; Wheldall & Mettem, 1985. Kalkowski (1995) notes that research in other subjects such as science, social studies and health is not as abundant but still supports positive achievement outcomes.

Results across various levels of education are as consistent as they are across disciplines. Several studies have suggested cognitive benefits for primary students (Allen & Feldman, 1973; Devin-Sheehan, Feldman & Allen, 1976; Gartner, Kohler & Riessman, 1971; Greenwood, 1991; Greenwood, Carta, & Kamps, 1990; Greenwood, Delquardi, & Hall, 1989), secondary students (Byrd, 1990; Maheady, Sacca, & Harper, 1988) and college students (Goldschmid & Goldschmid, 1976; Leelawong et. al., 2002. Indeed, even meta-analyses have supported the implementation of peer tutoring across subjects and levels (Britz, Dixon, & McLaughlin, 1989; Byrd, 1990; Cohen & Kulik, 1981; Cohen, Kulik, & Kulik, 1982).
The most cited reasons for the success of peer and cross-age tutoring involve the advantages of communication between people in roughly the same zone of proximal development. Tasks within the zone of proximal development are tasks that a student cannot do alone but can do with assistance from a more capable peer or adult (Vygotsky, 1978). Some have argued that an advantage of peer tutoring is that both tutors and tutees speak a common language. (Cazden, 1986; Hedin, 1987; KalKowski, 1995). Damon and Phelps (1989) suggest that the equality of status between tutor and tutee facilitates a more productive discourse, diminishing the passivity common in the adult-child instructional relationship.

Beyond the advantages of a common lexicon, there are likely other factors that contribute to the success of peer tutoring. Schunn (2004) describes the paradox of the expert, referring to the expert’s sometimes inability to deconstruct processes that for him/her have become automatic. As a process becomes more practiced and more automatic it becomes less inspectible, therefore placing the expert in a questionable position to explain the process to a novice. For example, a guitarist that has mastered a piece may be hard-pressed to articulate what his fingers are doing. Cohen (1986) has speculated that the shared cognitive framework enables peer tutors to present material in terms their tutees understand.

A recent study supports the notion that novices are in a better position to guide their peers. Students in a 12-week summer class at the University of Pittsburgh were assigned to write two drafts of a paper, using feedback from the first draft to make improvements on the second draft. Students were assigned to three groups, each with a
different source of feedback on the first draft. The sources of feedback were: (a) single expert, (b) single peer, and (c) multiple peers. The study showed increasing writing quality across these treatments, with the single expert group showing the least improvement and the multiple peer group showing the most improvement. There was a significant difference between the single peer and multiple peer groups. (Cho & Schunn, 2007).

Other researchers have cited advantages of peer teaching. Allen and Feldman (1976) discovered that third and sixth grade students were more accurate than experienced teachers in determining from nonverbal behavior whether peers understood lessons. A high school art teacher found that her students were more specific in writing criteria to prepare one another for an assignment than she would have written herself (Milbrandt, Felts, Richards, & Abghari, 2004). Others have pointed to the likely advantages in modeling when the model and learner are similar (Cohen, 1986). Several researchers have suggested that achievement gains are more pronounced in the tutor than the tutee (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Gaustad, 1992; Graesser, Person, & Magliano, 1995; Johnson, Sulzer-Azaroff, & Maass, 1977, Morgan & Toy, 1970). Some students attribute their gains to the responsibility of teaching, while others suggest they are due to clearly organizing the conceptual materials. (Biswas, Schwartz, Bransford, & TAG-V, 2001). Such claims are consistent with the articulation literature, suggesting the process of explaining involves a cognitive organization that may deepen understanding.
Resonant of Weigel’s (2003) assertion that technological development is providing new ways for teach to learn approaches to thrive, some researchers have begun to develop and study the use of computer-based teachable agents. A teachable agent is an intelligent computer character designed to learn information students impart through concept maps, which help to “categorize groups of objects and express interactions among them” (p. 246, Leelawong et al., 2002). Students can query and quiz the teachable agent to check for accuracy in their conceptual representations. Students are asked to analyze the responses they get from agents and the additional feedback from a teaching expert. The process helps students to identify problems and strengths in their concept maps. Research has shown that quizzing with pre-scripted basic questions and querying with student-designed reasoning questions produce different, but complementary outcomes. While quizzing reduces irrelevant information and increases causal information in student concept maps, querying increases interconnectedness of concepts (Leelawong et al, 2002). The authors deduce the importance of including a variety of feedback forms in this teaching process. Feedback from the teachable agent is designed to emulate feedback from students during teaching.

Leelawong et al. point out that teachers’ knowledge structures are organized as much by answering student questions and providing feedback as by preparing materials to teach (2002). While this particular application of teachable agents is known to work well in scientific domains due to clearly identifiable knowledge hierarchies and cause-effect conceptual relationships, it is likely, though unproven, that some of these principles are generalizeable to other domains. For example, are there similar meta-cognitive
advantages in asking students to provide feedback to peers on their writing? Would providing feedback on various dimensions of peer writing improve one’s own performance on those dimensions? Such questions have not been answered definitively in the literature, and underlie the purpose of this dissertation. Peer review literature and some of the related rationale to the design of the online peer review system employed in the current study are detailed in the following section.

Measuring Writing Quality

The practice of grading is inexorable, yet is often averted in academic discourse (Zak & Weaver, 1998). It is difficult to venture into a discussion about writing assessment, evaluation, or grading without encountering larger, more general questions about the meaning and purpose of grades. What is the purpose of grading? Can a singular symbol convey the complexities behind it? (Yancey & Huot, 1998). Are faculty members consciously or unconsciously protecting the power stream that springs from grading? (Zak & Weaver, 1998). Yet these and other questions seem amplified when it comes to composition. Writing is so complex, contextual, and interpretive that its evaluative form eludes standardization. Boyd (1998) quotes Harvard Professor Charles Copeland who observed: “grades that stand for an instructor’s impression of a piece of writing cannot be mathematically precise” (6). Years later, William McColly (1970) would voice a similar, yet broader observation: “There is a dramatic disagreement among the hundreds of thousands of English teachers as to what good writing is and is not, and also, many English teachers cannot be objective in judging student writing or in reporting their judgments” (148).
Students, keenly aware of the evaluative latitude given to graders of writing, often perceive the process as a game of figuring out what instructors want and giving it to them (Bernard-Donals, 1998). Evaluation practices that attempt to subvert or hide the game are ineffective and therefore met with suspicion by students and effectively underscore the importance of grades. Bernard-Donals writes:

But even evaluating and liking don't divert our very intelligent students from knowing even safe grading practices, inserted into a culture characterized by competition and hierarchy, are something of a ruse. In other words, our students are better readers of culture than we give them credit for, and when they see us try to do an end-run around grades, they rightfully chuckle out loud, but then quietly decide for themselves how to face an institution we won't. (64)

Implied in this observation is the notion that students might actually welcome the inevitability of grade acquisition because it is familiar, and therefore conquerable, territory. Students become skilled at grade acquisition because they get practice every time they take a class. It is conceivable that for many students, other skills might function subordinately to the overarching skill of grade control.

Even teachers, when placed in the role of students, suddenly tout the virtues of letter grades. At a summer writing institute in 1994, Anderson and Speck (1997), who expected participants (writing teachers) to be sensitive to the arbitrary nature of grades were surprised at participants' overwhelming advocacy for letter grades, even after receiving significant feedback in other, purportedly more meaningful, forms. Some participants even argued that grades represented an objective evaluation of writing.
Responses to a questionnaire at the end of the institute indicated that the majority of participants believed that grades help people improve future performance (1997). These data suggest that faculty might give attention to implications of a grade-centric culture as well as examine the ways their opinions are informed by the power of their positions. Faculty may want to more closely observe the experience of students, as well as their own biases, in order to better understand the dynamics of the grade game.

While the literature on assessment is brimming with theoretical debate about the purposes and processes of writing evaluation, there seems to be almost no discussion about specific criteria that should or could guide writing assessment. Even literature with the word *grading* in its title rarely, if ever, includes advice about the mechanics of determining grades (Anderson & Speck, 1997). Perhaps the absence of specific grading-process recommendations can be attributed to wide acceptance of the notion, advocated most prominently by Huot (1990, 1996, 2002), that writing evaluation criteria should be locally and contextually determined. In other words, it may be that writing is simply too subjective and contextual for standard practices to emerge. Working from this perspective, writing quality must be defined individually at least course by course, and more likely assignment by assignment. Boyd (1998) argues that attempts to standardize grading have always resulted in a focus on low-level aspects of writing, such as an over-emphasis on mechanics. Yancey and Huot (1998) propose that the goal of writing evaluation should be to move students to develop their own standards. Further, they suggest that writing quality cannot be quantified and should focus more on a description of process and potential within a context.
One of the stronger voices in philosophical conversations about grading composition is that of Peter Elbow (1998), who asserts, “Grading is not inevitable . . . institutions can prosper without grades” (p. 171). Knowing the low likelihood that very many institutions could or would embrace the notion of eliminating grades altogether, he proceeds to argue for at least minimal grading. He describes a process he calls “liking,” which is similar to what most compositionists call “responding,” and claims that it should be separated from evaluation. Regarding grading, his primary argument is for explicit criteria that allow for a distribution of attributes and acknowledge contextuality of writing. He also suggests that instructors use grids to quantify evaluation across multiple dimensions (1998). Elbow believes these shorter comments are more useful to students. The criteria in Elbow’s grid would have been communicated with students prior to the writing process. A drastic extension of this criterion-referenced method is contract grading, where students know exactly what they must do to attain various grades rather than being graded upon the qualitative merits of their work as judged by a teacher post hoc. While some of Elbow’s ideas about writing evaluation might be characterized as radical, he subtly acknowledges the cultural and functional inevitability of grades.

Other participants in the composition-grading dialogue recognize practical and cultural limitations to the adoption of alternative grading philosophies and practices. Bernard-Donals (1998), for example, points out that contract grading invites student contestation of grading criteria, because there is no price for such contestation, and contract grading encourages students to strongly guard their possibly misguided positions. In reply to those who urge a departure from evaluation in writing, Haswell
(1999) takes issue with the position that evaluation is not part of an authentic act of writing. He points out that all response is “situated and value-laden” (p. 286) and that writers usually do and should want to know if a current piece is better than previous work and how its qualities are perceived by an audience.

Some of the literature emphasizing practical approaches to evaluation has identified contradictions between philosophies of grading and practice (Maylath, 1998). For example, writing instructors teach process but still grade products (Appleman & Green, 1993). Even as early as the 1960’s researchers were reporting a “halo effect,” where teachers were said to assign undue merit to papers that exhibited only mechanical correctness (Scannell & Marshall, 1966). Teachers were much more concerned with mechanics than they purported or hoped to be. A similar study in 1977 showed an inverse relationship between the criteria some high school English teachers sought to use and what they actually employed (Harris, 1977). Other investigations have shown that sets of student essays received higher grades when readers believed they were from honors students versus average students (Diederich, 1965; Rigsby, 1987). One researcher who taught simultaneously at two different institutions learned that she applied grading practices inconsistently across the two schools, recognizing that she had been influenced by the schools different demographics (Uchmanowicz, 1998).

In review of two books on writing assessment, Haswell (1999) argues that the current obsession with credits is not a result of grading, but rather a result of reducing “courses of study” to “courses” (p.288, 1999), an acknowledgement of the commoditization of academic credits. Credit hours have been "standardized for exchange
even more than grades have" (p. 289) and issues of course are being confounded with
issues of grades. Citing Sternglass' longitudinal study (1997), which suggests that
learning occurs across courses, Haswell notes, "It is the autonomous course that relieves
us of the patience and time needed . . . to know how students really learn"(p. 294, 1999).
Implicit in Haswell’s statement is a suggestion that over-emphasis on individualization
and contextualization might lead to the counter-productive obscuring of cross-course
connections. Haswell tacitly makes the case that, not only should we accept the
inevitability of evaluation, but we might also want to look for ways to approach it more
institutionally.

Yancey (1999) identifies three overlapping waves of writing assessment: The
first, from roughly 1950-1970, is characterized by objective tests and indirect measures,
such as multiple choice tests measuring the use of grammar and syntax. The second
wave, from roughly 1970-1986, is characterized by the holistically-scored essay, often
called “direct assessment.” The third wave, from 1986- present, is characterized by
portfolio assessment and programmatic assessment. Yancey describes how direct
assessment began with one draft, led to two drafts, then to two drafts plus authorial
commentary, then to multiple drafts with reflection (portfolio). In the first wave, the
emphasis was on reliability. In the second wave, the emphasis shifted to validity, a result
of knowing more about writing. Finally, the third wave combines attention to both
validity and reliability (1999).

Bellamy (2006) characterizes the later developments in writing evaluation
practices somewhat differently than Yancey. Bellamy dates the shift to holistic grading
much earlier than Yancey, implying that it began at least prior to the 1960’s. This discrepancy alone might not be all that significant, as even Yancey points out, the waves are overlapping, non-discreet movements that have varying levels of durability across various contexts (1999). However, there is significant disagreement between Yancey and Bellamy about what has emerged to replace holistic grading. Rather than identifying a movement toward portfolio assessment, Bellamy sees trait-based assessment as the most salient current evaluation trend (2006). Interestingly, Bellamy also points to 1986 as a critical year, noting the appearance of Hillocks’ (1987) meta-analysis of twenty years of writing research. Revealed by the study, among many other findings, was that the use of scales—an application of trait-based evaluation—was one of the most effective practices in writing instruction (1987). Bellamy (2006) defines the use of scales as “the training of students to use sets of criteria with which to judge the quality of their work” (p. 2). He describes a six-trait scale that addresses the following criteria: ideas/content, organization and development, voice/tone/flavor, effective word choice, syntax, and writing conventions. A few studies have supported the use of this particular traits model. Arter, Spandel, Culham, and Pollard (1994) showed significant gains in fifth and sixth graders provided training in three of the traits. Coe (as cited in Bellamy, 2006) found that all six of the traits were strongly predictive of success on the Washington Assessment of Student Learning in Writing, especially in the subscales of ideas, conventions, and sentence fluency.

While Yancey and Bellamy described the ways evaluation has evolved specifically in the field of composition, the evolution has bearing on the practice of
WAC, because composition instructors are often the designers and implementers of WAC professional development activities and training (Bazerman et al., 2005). In essence, developments in writing research and educational psychology have led to the widespread adoption of the following principles: (a) iterative drafting practices which emphasis knowledge transformation rather than knowledge telling, (b) inclusion of meta-cognitive activities aimed at providing students with a stronger grasp of their own rhetorical processes, strengths, and deficiencies, (c) contextualized writing assignments, (d) formative evaluation practices (e) attention to validity and reliability (f) analytic, trait-based evaluation. These now fundamental principles are easily realized within a WAC paradigm, which emphasizes longitudinal development of writing skills across several courses, assignments and forms. Such principles are even more tenable with powerful new web applications for instructor and peer evaluation of student writing.

One such system, called ICON (for Interactive Composition Online) (Wasley, 2006), has been adopted by the freshman composition program at Texas Tech University. The system, designed to facilitate the teaching of composition to 3000 freshman by inexperienced graduate students—a challenge common to large research institutions—targets several of the principles described above. The system is designed to compel graduate students to follow common, trait-based evaluation procedures in grading freshman writing assignments. Grading quotas are built into the conditions of graduate students’ assistantships. Due to efficiencies in the online system, proponents claim that students are able to write more frequently relative to traditional treatments of writing instruction. Among other reported advantages, reliability is believed to be enhanced
through a process of averaging student scores from multiple raters. To date there have been no data collected to corroborate proponents’ claims, yet with its nine linked databases and “201 discrete, searchable/Sortable chunks of information” (p. A6, Wasley, 2006) the system promises to become well-studied.

Peer Review

What if there was a way to provide a consistent structure to grading, while at the same time allowing for individualization and contextualization? What if there was a way to engage students in the cognitively productive act of evaluation, while at the same time providing a valid and reliable measure of student writing quality? What if there was a way to open up the possibility of writing within discipline classes while mediating the biases in traditional grading practices? Online peer review systems have advanced to the point that they have opened the door to all of these seemingly lofty and sometimes incompatible demands. As in any evolution, moving peer review into online delivery has built on a foundation of traditional peer review while developing in its new dimensions. Online peer review systems address many of the current concerns about implementing writing across the curriculum, especially evaluation.

Peer review, whether traditional or online, is known by different names and is practiced in different flavors. Peer assessment, peer response, peer critique, peer evaluation, peer grading, and peer revision are all terms that are sometimes used interchangeably. Usually the choice of label depends on the extent to which student grades are determined by the process and also the extent to which students advise one another during revision. Finite definitions of these terms are not available because their
characteristics are not necessarily discreet. For example, a system might provide high levels of input on student grades while also providing extensive opportunity for student to student advising. For the purpose of consistency, peer review will be used here to describe the family of activities that place students in the reciprocal relationship of providing and receiving feedback on their writing.

Intuitively, peer review activities seem to offer several advantages over traditional means of evaluating and responding to student writing. Perhaps the most cited advantage of peer review is its efficiency (Cho & Schunn, 2007; Cho, Schunn, & Wilson, 2006, 1995; Gerson, 1988; Kugel, 1989; Marcoulides & Simkin, 1991). Others have noted that peer review involves students more actively in their learning by giving them exposure to and practice in evaluation of writing, thereby raising their awareness of their own rhetorical strengths and weaknesses (Haaga, 1993; Hu, 2005; Marcoulides & Simkin, 1991, 1995; Paulus, 1999; Tsui & Ng, 2000). Another argument for peer review stresses the possibility that students might actually provide more useful feedback to one another than instructors because of students’ cognitive proximity to their peers (Cho & Schunn, 2007; Hu, 2005). Finally, some have argued that peer review honors current models of writing process that emphasize recursive, social construction of knowledge (Bereiter & Scardamalia, 1987; Hu, 2005; Marcoulides & Simkin, 1995). Thus, student peer review mimics the form of professional inquiry and is thought to prepare students for professional, academic, and corporate life (Berg, 1999; Marcoulides & Simkin, 1995; Sullivan, Brown, & Nielson, 1998). Gillam (1990) suggests that by communicating their reviews to fellow writers, students develop a writing vocabulary.
Yet, despite strong intuitive support for peer review, as well as widespread adoption in composition classes, it has received little attention from empirical researchers. Haswell (2005) recently noted that peer review of undergraduate writing is the “least studied” of evaluation practices (p. 211). While there are many scholarly articles on the use of peer review to support writing, most provide anecdotal accounts and observations about its utility (Dossin, 2003; Marcellus, 2001; Pitts, 1988; Shaw, 2002) or are otherwise descriptive in nature (DiGiovanni & Nagaswami, 2001; Hu, 2005). A recent review by Cho, Wilson, and Schunn (2006) turned up only six previous studies that had examined validity or reliability of peer assessments of writing (Cheng & Warren, 1999; Falchikov, 1986; Haaga, 1993; Marcoulides & Simkin, 1995; Mowl & Pain, 1995; Stefani, 1994). Cho, Wilson, and Schunn (2006) note that the previous literature shows mixed results and either measured validity or reliability, but never both, making it difficult to attribute problems to one or the other. Of the previous studies, Marcoulides and Simkin’s (1995) revealed the most compelling evidence of peer review reliability. Using a percent variance analysis approach, they found that 0% of the variability of scores could be attributed to peer reviewer effects. The authors noted, however, that student reasoning behind their scores was not near as consistent as the actual scores (1995). Though compelling, the results of this study have limited generalizability and have no bearing at all on questions of validity.

Believing in the potential of online peer review to facilitate writing in disciplines other than composition, Cho and Schunn designed and developed SWoRD, a web-based peer review system that uses advanced algorithms to compute student writing and
reviewing scores, primarily through the aggregation and weighting of peer evaluations (Cho & Schunn, 2007). SWoRD is an acronym for Scaffolded Writing and ReWriting in the Discipline, and the system is described in detail in the methods section of this dissertation. Seeking to address shortcomings in peer review literature and to validate SWoRD, Cho, Wilson, and Schunn (2006) designed a study to examine the validity and reliability of scores generated via online peer review. In particular, the investigation was designed to assess validity and reliability across several different courses, disciplines, and levels with attention toward discerning appropriate numbers of reviewers (Cho, Wilson, & Schunn, 2006). Another distinction of the investigation is its consideration of disparate notions of validity and reliability between instructors and students. Cho, Wilson, and Schunn collected data from 16 different courses across several disciplines at the graduate and undergraduate level from four different universities. Reliability was assessed in all 16 courses, but validity was only assessed in five of the courses and was defined as a correlation between mean peer ratings and instructor ratings. Results showed that SWoRD scores were highly reliable and valid, regardless of level or discipline, when at least four reviewer scores were aggregated. In two of the courses, a writing expert’s external evaluations showed lower correlation to instructor ratings than SWoRD scores.

The data from Cho, Wilson, and Schunn’s (2006) study reveals how computer-mediation can improve the reliability and validity of peer review activities while simultaneously improving their functionality. SWoRD is one of a handful of computer-mediated systems that have emerged in recent years. Similar systems include: Calibrated Peer Review or CPR (Robinson, 2001), Norton Textra Connect (DiGiovanni &
Nagaswami, 2001), Daedalus Integrated Writing Environment (Daedalus Group, Inc., 2007), and Interactive Composition Online (Wasley, 2006). Of these programs, the only two programs specifically offering online peer review are SWoRD and CPR. The two programs are similar, but the SWoRD process involves two drafts, where CPR only involves a single draft. Also, CPR focuses more attention on calibration and includes student self-assessment (Cho & Schunn, 2007; Robinson, 2001). A quasi-experimental study comparing the methods of faculty review, group review, and CPR across 148 undergraduate students in naturally-formed sections of a business communications class revealed no significant differences among the teaching and evaluation methods, supporting the notion that CPR is at least as useful as the other methods.

Practitioners have touted the benefits of computer-mediated peer review since the late 1990’s (Sullivan, Brown, & Nielson, 1998). Available options at that time paled in comparison to robust web applications available at the time of this writing. Online peer review solves several of the challenges and problems with face-to-face peer review processes. One significant advantage of online review is its ability to facilitate anonymity, freeing the process from some of the problematic interpersonal dynamics and biases that plague traditional peer review (MacLeod, 1999; Wasley, 2006). One of the most prominent advantages of online peer review is its efficiency. With computer workflows, several aspects of the peer review process are improved: submission of papers, paper distribution (Palmquist, 1993; Sullivan, Brown, & Nielson, 1998), score calculation and weighting, application of complex grading algorithms (Cho & Schunn, 2007; Robinson, 2001), and provision of feedback to participants (Matsumura & Hann,
2004; Tuzi, 2004). In fact, efficiencies between traditional and online peer review systems are so vastly different that much of the capacity of online peer review systems is practically impossible in face-to-face applications. For example, the time required for instructors to individually and manually calculate scores and aggregations effectively prevents instructors from keeping pace with assignments, even if they are committed to doing so. It has also been noted that online systems allow instructors to better monitor peer review processes and provide guidance to those who need it (DiGiovanni & Nagaswami, 2001). Finally, robust online peer review systems address one of the most common objections to peer review: the questionable ability of students to accurately review their peers. SWoRD in particular has been proven to adequately scaffold the weaknesses of individual student ratings through its aggregation and weighting of student scores (Cho, Wilson, & Schunn, 2006).

Clearly there are compelling reasons to believe that online peer review has the potential to rejuvenate writing in the disciplines while solving many of the persistent problems that have been associated with traditional face to face peer review activities. However, such systems are too young to have been fully explored in the literature, and several questions require investigation:

1. What features maximize the validity and reliability of student scores?
2. What design characteristics have the largest impact on learning?
3. What aspects of this new peer review process are most critical to developing writers?
This dissertation seeks to validate the tacit assumption that reviewing activities lead to improved writing skills for the reviewer. With this focus on the reviewer, what particular reviewing behaviors are most beneficial? For example, will the reviewer see more benefits from producing elaborate comments during feedback, or terser, direct forms? Does the reviewer see benefits from exposure to prototypal models of helpful and unhelpful reviews? Hints about these issues may be found in the final sections of this literature review.

Feedback

Feedback was initially conceived by early educational psychologists as merely positive or negative reinforcement of behavior. Over the past several decades, instructional feedback has evolved into a sophisticated control model with the maturation of an information processing orientation to educational psychology. Information processing models, which form the foundation of cognitive psychology, treat the mind as a complex electrochemical computer that receives various inputs, makes appropriate manipulations and calculations, and outputs information. From this perspective, errors are not simply mistakes, but a source of information about cognitive processes. Relative to early operant models about research, information processing models provide stronger explanatory power, especially when it comes to feedback (Kulhavy & Stock, 1989).

The classic and most definitive feedback model, presented by Kulhavy and Stock (1989), proposes that feedback is composed of two components: verification and elaboration. Verification is simply an indication whether a student response is correct or incorrect. Elaboration consists of all other substantive information contained in a
feedback message. In the Kulhavy & Stock (1989) model, elaboration can be one of three types: (a) task specific (most often restatement of correct answer), (b) instruction-based (explanation of why a response was correct or not using lesson materials) or (c) extra-instructional (containing examples or analogies from outside the instruction). From the introduction of the model, response certitude (subjective certainty of respondent) emerged as a salient factor, predicting durability, or the “likelihood that an instructional response will be available for the learner’s use at some later point in time.” (p. 296, 1989). Based on their empirical studies, Kulhavy, et. al. (1990) propose that the ideal timing of feedback and degree of elaboration of such feedback is dependent upon one’s own rating of response confidence.

Proponents of this model herald its clear analysis of discreet feedback components, its presentation of testable terms, and its emphasis of the often-ignored internal processing of the feedback message. (Hancock, Thurman, & Hubbard, 1995). However, several objections to the Kulhavy/Stock model have been noted. These objections include:

1. The model doesn’t account for learners’ goals (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler, Winne, & McGinn, 1993; Dempsey, Driscoll, & Swindell, 1993; Hancock, Thurman, & Hubbard, 1993; Mory, 1992; Shutz, 1993)

2. The model assumes that processing is automatic for all subjects, without regard for how mindful they may or may not be (Bangert-Drowns, Kulik,
Kulik, & Morgan, 1991; Hancock, Hubbard, & Thurman, 1992; Swindell & Walls, 1993)

3. The model doesn’t account for individual learner characteristics (Butler, Winne, and McGinn, 1993; Dempsey, Driscoll, & Swindell, 1993)

4. While the Kulhavy/Stock model does acknowledge the importance of learners’ perceptions, it still casts feedback as a stimulus that causes a learner to perform in a predictable manner (Hancock, Thurman, & Hubbard, 1995)

To address these concerns, Hancock, Thurman, & Hubbard (1995) expanded the control model to include consideration of learner goals, mindfulness, and other individual differences. Mason and Bruning (2001) have summarized the dimensions that may influence the effectiveness of feedback, including: elaboration, student achievement levels, depth of understanding, attitude toward feedback, learner control, response certitude, and timing.

From a practical standpoint, feedback that adapts to individual differences would best be handled by an advanced algorithm on a computer, as no human would be capable of calculating and rendering appropriate feedback within an adequate timeframe. Hannafin (1984) suggested, “one of the most powerful and important features of the computer is the virtually unlimited range of instructional control options available to designers” (p.6). Others have pointed out that computer-generated feedback can remain unbiased and accurate regardless of the student characteristics that might otherwise subjectively influence a person (Mason & Bruning, 2005). In fact, much of the literature
on feedback has been developed as part of research on computer-based instruction (CBI). While lending itself well to empirical testing, clarifying and addressing many of the relevant variables in traditional (multiple choice) assessment practices, the Kulhavy/Stock model may not be well-suited to more complex learning tasks, such as writing. Indeed Kulhavy and Stock (1989) acknowledged this limitation in their original publication on the feedback model: “We are aware that this particular framework may limit the model to small-step, recognition task demands. . .” (p. 290, 1989).

Considering the complex range of factors influencing feedback on simple recognitive task demands, it is no wonder that researchers of writing processes, with their almost infinite complexities, have not been able to apply the classic feedback model to their own examinations of feedback. Instead, research on feedback in writing has tended to focus on the type of feedback most useful to student writers, the degree to which such feedback leads to revisions, and student preferences regarding feedback. For example, in an elementary school setting, a process goal with progress feedback was found to be more beneficial to fifth-grade writing students than a product or general goal (Schunk and Swartz, 1993). All students in the study received performance feedback, but some students received additional feedback about their choice of processes and procedures. These results suggest that different types of feedback can have varied effects on student writers.

Another study, aimed at assessing the types of feedback offered by high school students during peer responding activities, revealed several categories of student response: global praise, personal response, text playback, sentence edits, word edits,
reader’s needs, and writer’s strategies. (Simmons, 2003). The study paired high school students with college writing classes across several institutions and demographics. Students tended to focus on surface level errors when responding to peer writing, leading the authors to conclude that it takes a great deal of time and practice to guide students to respond to writing on its deeper dimensions. Yet, even feedback from relatively inexperienced students tended to improve the writing of students who attended to their peers’ feedback (2003). Others studies have corroborated a few of these findings. For example, the tendency for students to focus on surface errors has been noted (Bridwell, 1980; Crowhurst, 1986; Dheram, 1995; Faigley & Witte, 1981). Also, several authors have pointed to the potential benefits of multiple reader/responders (Anson, 1999; Cho & Schunn, 2007; Kent, 2004; Matsumura & Hann, 2004).

As noted above, student preferences have also been the subject of feedback investigations. For example, through a multiple regression analysis involving 207 undergraduate Japanese students in English as Foreign Language writing classes, Matsumura and Hann (2004) showed that students’ choices of feedback method between face-to-face teacher feedback and online teacher and peer feedback varied as a function of the level of their computer anxiety. In addition, having the choice to use or not use computers helped both high and low anxiety students improve their writing. In another investigation of second-language writers, results suggest that students prefer face-to-face feedback conditions, but actually show more improvement from electronically facilitated feedback (Tuzi, 2004).
Most feedback and revision research has been directed at composition courses. Beason (1993) noted the lack of feedback and revision inquiry in non-composition courses and questioned the applicability of composition research in WAC contexts. His qualitative study of 20 students in four different courses examined following factors: (a) the criteria guiding student feedback, (b) the extent students used feedback, (c) the criteria most affected by the revisions, and (d) the extent of the revisions. Results show that advising, praising, problem detecting, and editing were the most frequent aims of feedback. Responder comments, and the revisions to which they led, tended to focus on issues of expression, development and support. Students paid more attention to teacher comments than peer comments in their revisions. Also, revisions were divided fairly equally between meaning-changing and non-meaning changing revisions. While this study helped to identify categories of response and revision that can guide future research, it offered no experimental controls and therefore its results are not easily generalized. Because patterns and frequencies of comments and revisions are described, the author calls the approach quantitative. However, the study is non-experimental and much closer to the naturalistic methodology dominating WAC literature than to the positivist leanings of empirical feedback research.

While a preference for interpretivist research might be expected, perhaps even desirable, within the domain of writing, there are many questions about feedback and revision in content classes that could be well-served by experimental research. Pragmatically, instructors in disciplines outside of writing, where writing activities can easily become unwieldy, might benefit most from specific guidance regarding known
variables. For example, what specific types of feedback are known to generalize? What effects can be observed on the feedback provider in peer review processes? Do different levels of elaboration by the feedback provider lead to improved writing quality by the feedback provider?

Perhaps writing research can make use of the Kulhavy/Stock model after all, by separately examining the components of the complex act of writing. For example, because research on feedback tends to favor feedback elaboration more than not, it seems likely that encouraging more elaborate feedback during peer review might lead to improvements in student writing (Hancock, Thurman, & Hubbard, 1995; Mason & Bruning 2001; Pridemore & Klein, 1991). Considering the research on articulation and self-explaining, there is reason to believe that feedback elaboration treatments stand to benefit responders at least as much as writers under review.

Models and Modeling

Margolis’ (1987) suggests that all thinking and judgment can be reduced to pattern recognition. If so, it is likely that learning in many, if not all domains, can be facilitated through the use of examples. A music student learning the distinguishing features of the baroque style undoubtedly listens to examples. Even a music student learning to play in the baroque style most likely listens to examples, perhaps even examples that push or stretch the boundaries of the style. An economics student learning about supply and demand might be given an example of hurricanes destroying orange crops in Florida and subsequent increases in the price of orange juice. Certainly the ways that examples are used and studied vary from discipline to discipline. Yet, despite its
heterogeneity, literature related to the instructional use of examples reveals a clear
distinction between *models* and *modeling*. Models employ the use of products in
demonstration, while modeling involves the examination of a process at work. In other
words, models might be worked-out examples, where modeling might involve watching
an expert actually work through the solution to a problem.

Early examples research focused on concept learning and often measured a
student’s ability to properly identify concepts after being exposed to several examples
and non-examples (Atkinson, Derry, Renkl, & Wortham, 2000). These studies, executed
from the mid 1950’s through the 1970’s, were focused on the selection, presentation and
sequencing of examples (2000). Yet as early as the 1970’s researchers began to study the
use of examples in more complex forms of learning, including sophisticated
mathematical and scientific problem-solving. Sweller and Cooper (1985), for example,
revealed that students who were shown model problem solutions before solving complex
algebra problems demonstrated more efficient solution strategies and tended to focus on
the deeper structural aspects of problems more often attended to by experts (Cooper &
Sweller, 1987). These early laboratory experiments generated a new strand of classroom-
situated, “worked examples” (184) research (Atkinson et. al., 2000).

Yet, while research on the instructional use of examples continued to evolve and
flourish throughout the 1980’s and 1990’s, it was conducted mostly in the fields of
mathematics, the sciences and computer programming. While some have asserted that
worked examples research is especially relevant to instructional practice involving skills
acquisition, there have been few investigations into the effects of using examples in
teaching the skill of writing (Atkinson et. al., 2000), and the scarcity of research in the use of model texts is particularly pronounced (Charney & Carlson, 1995). Hillocks’ (1987) review of the previous twenty years of writing research reveals that, while the use of models had been studied in composition instruction, treatments tended to involve only surface feature identification, akin to declarative knowledge. For example, students might be shown samples of dangling participles before being asked to select them from a passage, and even this limited exploration ultimately produced mixed, if not discouraging results.

Charney and Carlson (1995) assert that the discrepancy between Hillocks’ (1987) mixed results and more consistent results in other domains (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Nitsch, 1977; Pirolli, 1991; Sweller & Cooper, 1985), might be explained by the fact that writing tasks are ill-defined relative to other problem-solving tasks that have been studied. Charney and Carlson (1995) also note that students often misuse models, “imitating their weaknesses as well as their strengths” (p. 90). Similarly, Smagorinsky (1992) observed that models were available before but not during student writing activities, underscoring the lack of consistently applied experimental treatments.

To address the gap in the literature regarding the effects of different kinds of models, Charney and Carlson (1995) examined how students make use of model texts as they learn to write within the conventions and restraints of a particular genre. They sought to examine whether writers benefitted from seeing models of varied quality levels with explicit information about their quality (1995). While the inclusion of varied quality
models or models with quality labels produced no effect, there was a main effect for using models in general. A holistic evaluation of student papers across four scales—inclusion of relevant information, exclusion of irrelevant information, organization, and elaboration—revealed significant differences between the models group and the no-models group only in the category of organization. Also, seeing a proposition in models, especially those of higher quality, increased the likelihood that students would use them in their own writings. The results of the study, especially regarding differences across various dimensions of writing, may explain previous mixed results. Clearly, more information is needed about the effects of models on student writing, and researchers should consider dimensions independently as well as holistically.

While use of model texts has received less attention in the 12 years since the Charney and Carlson study, an emerging body of literature is revealing more about modeling, the observation of another writer actively working through a writing task. This type of study requires writers to think aloud while performing a complex task. Researchers of modeling in writing tend to focus on the orchestration of writing processes, defined as “the temporal organization of cognitive activities” (Braaksma, Rijlaarsdam, van den Bergh, & van Hout-Wolters, 2004). Several studies have suggested that the timing of cognitive activities during the writing process has decisive impact on writing quality (Breetvelt, van den Bergh & Rijlaarsdam, 1994; van der Hoeven, 1997; van den Bergh & Rijlaarsdam, 1999; van den Bergh & Rijlaarsdam, 2001). In a study of 52 eighth-graders, students exposed to observational learning methods performed more meta-cognitive processes such as planning and goal orientation, showed a more varied
pattern of control activities, and ultimately produced higher quality texts (Braaksma, Rijlaarsdam, van den Bergh & van Hout-Wolters, 2004). Other researchers have shown positive effects of observational learning activities in writing instruction (Couzijn, 1999; Graham & Harris, 1994; Graham, Harris & Troia, 1998; Zimmerman & Kitsantas, 2002).

Yet, as compelling as it might be to study process over product, such investigations are often limited to laboratory settings. Refining the literature in ways that will be useful to learners and instructors requires some research in instructional settings. Observational learning activities, if performed in classes, are time consuming and logistically challenging, even in a class dedicated to the study of writing. It is unlikely that instructors across the curriculum will have the means or inclination to incorporate modeling in writing activities. Further, what has been learned from modeling research might be similar to what has been learned from model research. For example, the impacts on structure noted by Charney and Carlson (1995) are more strongly associated with the meta-cognitive control processes occurring during writing—the processes Bereiter and Scardamalia (1987) positioned in the rhetorical problem space of their model. This observation is consistent with the findings of Braaksma et al. (2002, 2004) showing clear impacts on planning and other orchestration activities during writing. In each case, effects are more obvious in the rhetorical problem space than in the content problem space. In other words, observers of models or modeling in writing seem to absorb methods of rhetorical control rather than content insight.

Certainly models and modeling appear to have impact on the activities of writers, but much more refinement is needed as to the specific sub-processes affected and the
most suitable instructional applications for maximizing their benefits. In particular, little is known about how the use of examples can best be incorporated into the writing activities of non-composition classes. For example, do models of good and poor feedback help students new to reviewing become better reviewers in WAC contexts? If students are provided models of writing reviews and subsequently apply them to their reviews, will the review process ultimately translate into stronger writing skills? It seems likely that the act of reviewing itself might be beneficial to students, giving them exposure to various models of writing. Also, the quality of the reviewing experience might be enhanced and more deeply learned by students exposed to examples of good and poor reviews.

Conclusion

If, as Bereiter and Scardamalia (1987) suggest, writing has the power to be knowledge transforming, it represents an avenue for learning in any field, any domain. Of course, as post-process theorists and WID advocates remind us, writing is highly contextual (Bazerman et. al., 2005). Still, much more is common about the act of writing than is contextual. Besides a common language and conventions, the mental processes of writers have proven to be consistent. Nearly 200 studies, which were the foundation of Bereiter and Scardamalia’s (1987) model, show consistent patterns in the processes of novice and expert writers. Certainly those patterns likely flex and fluctuate relative to the context, especially with regard to an ever-changing public discourse, but there are consistent elements that can and should be better understood in order to foster improved facilitation of writing as a learning tool across disciplines.
As supported in the literature, online peer review is an approach to writing that addresses several of the issues associated with writing outside of composition classrooms. For example, the ability to count on the validity and reliability of composite student-generated scores frees non-composition faculty from the burden of grading student papers above and beyond the many other discipline-specific activities they must manage in their classes. Because papers are managed by an online system, they do not take in-class time or face any of the disadvantages of in-class peer-review activities, such as student-to-student biases or students being less-than-honest in their critiques of peers who they like or don’t like personally. Online peer review systems can expedite several cumbersome processes that would otherwise prohibit writing activities in the disciplines, such as: (a) submittal of papers, (b) distribution of papers for reviewing, (c) calculation of composite student scores and weighting, and (d) timely distribution of feedback. And, because online peer review systems store their review data in easily accessible databases, they also facilitate research on writing.

By supporting students’ reviews of peer papers, the SWoRD system encourages meta-cognitive mental activities, especially level three verbalizations, which, based on articulation literature and cognitive psychology literature, seem likely to help students organize their own cognitive processes as they write. Yet, so far most studies of self-explaining and other forms of articulation have been conducted with students working through science and math problems. More information is needed about the role of articulation in the verbal problem-solving activities embedded in writing activities. This
dissertation examines the effects of reviewing on subsequent writing of the reviewer. In particular, how does writing improve as a result of reviewing activities?

Of course there are many factors that may impact the effects of reviewing activities. For example, because reviewing is less familiar to students, its use evokes questions related to cognitive load. Cognitive load theory, in very simple terms, suggests that working memory limitations affect the amount of new information that can be digested at one time, particularly regarding complex tasks requiring processing of any kind (van Merriënboer & Ayres, 2005). On one hand, it seems likely that asking students to provide more elaborate feedback to their peers will result in increased cognitive organization. On the other hand, such elaboration probably introduces increased cognitive load and has the potential to overwhelm students. Still, because feedback research tends to favor elaboration, it seems more likely that students will benefit from providing more elaborate feedback on peer writing. Another factor that might be related to cognitive load and also might impact the effects of reviewing activities is the use of examples of good or bad feedback. Examples research is generally supportive of the use of examples and modeling across several instructional contexts, and therefore would suggest that providing examples of good and poor feedback will lead to improved cognitive organization during reviewing activities that might translate to a student’s subsequent writing. However, not only has the treatment of examples in this particular context been unexamined, it is possible that combining the use of examples in addition to more elaborate feedback treatments may lead to counterproductive levels of cognitive load. Therefore, the current investigation will measure the main effects of a feedback
elaboration treatment and an examples treatment, and it will also measure interaction between these two treatments.
CHAPTER III

METHODS

This chapter presents a description of the research purpose and questions, participants, treatment groups, apparatus, instrumentation, procedures, research hypotheses, design, and data analysis. The purpose of this dissertation is to examine the effects of reviewing peer papers on the subsequent writing of the reviewer within an online, asynchronous peer review system. The study addresses the following research questions in two phases:

Phase I

1. Does the articulation of feedback in a web-based peer review system result in higher quality of reviewers’ subsequent writing?

Phase II

1. Does the provision of examples of helpful and unhelpful reviews result in higher quality of reviewers’ subsequent writing?
2. Does more elaborate feedback by reviewers result in higher quality of reviewers’ subsequent writing?
3. Is there an interaction between example exposure type and feedback elaboration type?

Participants

Participants include 114 students selected from 10 sections of a sophomore-level Educational Psychology course at a large, Midwestern state university. The 10 sections of the course are distributed across four different instructors. Since the course is required of...
all education majors, the sample is likely representative of the population of education majors at this institution, barring any unusual historical factors. The course is usually taken by all education majors during the second year of a four-year program. Participants range in age from 19-44. Males composed 46% of the sample (n=53), and females composed 54% of the sample (n=61). Grade-Point-Average at the time of the study ranged from .75-4.0, with mean 3.21 and median 3.27.

Apparatus

SWoRD (Scaffolded Writing and Rewriting in the Discipline) is an asynchronous web-based reciprocal peer review system designed to facilitate student peer review of writing in content classes of any field. The system manages the submission of student papers and their subsequent distribution to a set of peer reviewers. Students using the system assume pseudonyms in order to maintain anonymity. For each assignment, students are required to engage in writing and reviewing activities across two drafts of a paper. After each draft submittal deadline has passed, papers are randomly assigned to 5-6 peer reviewers. Therefore, writers become reviewers of peer papers, and each student’s paper is scored through the aggregation of peer scores with an internal proprietary algorithm. All students are graded on both their writing and reviewing. As reviewers, students are guided to evaluate writing on three dimensions: flow, logic, and insight. Each of these dimensions is rated by reviewers on a seven-point scale. Reviews are weighted based on three accuracy indices: systematic differences, consistency, and spread. Reviewers are also back-reviewed by writers they have reviewed, giving the reviewer an indication of how helpful the review was to the writer. Total review scores are divided
evenly between review accuracy and back-review, which is a rating of helpfulness given to reviewers by writers.

Participants were given directions through both a written instructional pack and verbal direction from the instructor (see Appendices B-I). These printed materials supplemented the online directions provided by the SWoRD system. The additional directions were provided as part of the experimental variations and to situate the assignment within the context of this particular course. Student review activities required access to a computer with Internet connectivity, a web browser, a word processing program capable of generating a Microsoft Word document (.doc), a Rich Text Format (.rtf) document, or a Portable Document Format (.pdf). The system is designed to guide novices through the process of submittal, and therefore does not require any special skills beyond the use of standard web browsers and simple web-based forms.

Instrumentation

Data for this study were collected from two sources: (a) peer-generated writing scores as aggregated and reported by the SWoRD system and (b) averages of three expert scores derived from the same criteria and dimensions. As described above, SWoRD is an asynchronous, web-based peer review system for distributing and evaluating student writing. It performs a proprietary algorithm to student scores, consisting of averaging and weighting multiple peer ratings to obtain accurate writing and review scores. The seven-point Likert scale used by students to review peer papers on three separate dimensions of writing is consistent with common trait-based evaluation methods (Bellamy, 2006).
Because SWoRD creates a composite writing score from the three dimensions of writing it evaluates, both composite and individual dimension scores are available for analysis.

Three experts were asked to evaluate student papers on the same criteria and with the same Likert scale used by student reviewers during the SWoRD process. Each of the experts have several years of experience in evaluating student writing in an academic setting, either at the secondary or university level. Each expert also has a minimum of a master’s degree in either education or in English. Experts were asked to perform blind evaluations, by evaluating papers independently with no evidence of previous evaluations. Experts were given the same rubrics and assignment descriptions given to student participants.

The SWoRD system is appropriate for this study for several reasons. The systematic weighting of student reviews helps to mediate some of the common concerns associated with peer review, including internal and external inconsistency among peer reviewers. Further, the system requires students to evaluate writing across three dimensions using a seven-point rubric, providing reviewers a guided range of multi-dimensional evaluation. This approach honors the complexity of the writing process without becoming unwieldy.

The system’s web-based design offers several advantages. First, by introducing anonymity to participants, it eliminates the potential distractions of personality conflicts sometimes encountered in face-to-face peer review. Second, the anonymity also relieves pressures to favor the work of friends. Random assignment of reviewers to artifacts eliminates the biases that might come from self-selection. Finally, since the system is
programmed to make all of the calculations of reviewer and writing scores automatically, it makes feasible a process that might otherwise be prohibitively cumbersome.

SWoRD-generated scores have proven to have high validity and reliability through an investigation spanning several different courses, disciplines, and levels (Cho, Wilson, and Schunn, 2006). The study included 16 different courses across several disciplines at the graduate and undergraduate level from four different universities. Data were collected from 708 students. Reliability was assessed in all 16 courses, but validity was only assessed in five of the courses and was defined as a correlation between mean peer ratings and instructor ratings. The authors note that validity is defined differently by students than by instructors, and therefore measure both of these perspectives independently. They use instructor scores as the reference for peer-score validity, since instructor scores are often the only measure of student writing quality and instructors have the authority and responsibility to determine appropriate assessments of writing quality in their courses. SWoRD scores had high correlations between peer rating and instructor ratings (Cho, Wilson, and Schunn, 2006). Since each student reviews only a small portion of all student papers, and that subset is different from student to student, McGraw and Wong’s (1996) Intra-class correlations (ICC) were used to assess SWoRD’s score reliability. Cho, Schunn, and Wilson (2006) call this measure effective reliability. In their study, SWoRD scores were proven to be highly valid and reliable.

Given these reliability and validity results, it is reasonable to expect SWoRD scores to provide valid and reliable measures of writing, however to control for the possibility that the current context may have influenced the reliability and validity of
SWoRD scores, these scores were compared, through simple Pearson correlations, to the average scores of three expert raters with experience grading student writing. As shown in Table 1, results indicate that these scores are correlated. However, the relative spread of expert versus SWoRD scores (see Figures 1-2, Chapter V) indicates that experts made more distinctions in their ratings of paper quality.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow</td>
<td>.329</td>
<td>.000</td>
</tr>
<tr>
<td>Logic</td>
<td>.373</td>
<td>.000</td>
</tr>
<tr>
<td>Insight</td>
<td>.286</td>
<td>.002</td>
</tr>
</tbody>
</table>

A complete discussion of this issue will be reserved for Chapter V, but given constraints of the research environment, averaged expert scores are considered the primary source of data for this dissertation. SWoRD scores appear in the Appendices. There are two reasons why expert scores are a better choice here: (a) the timeframe available for the assignment prevented use of an optional calibration step that may have improved peer-reviewing accuracy, and (b) explanations provided to participants about the SWoRD scoring mechanisms may have inadvertently encouraged narrower ranges of ratings by reviewers. This study is focused on effects of review processes on student writing, not on validity.
and reliability of SWoRD scores, which have been well-established in several contexts (Cho, Wilson, and Schunn, 2006).

IntraClass Correlations (Shrout & Fleiss, 1979) were used to establish inter-rater reliability of expert ratings. IntraClass correlations (ICC) are a common measure of reliability when multiple judges are used to assess a common set of targets or subjects. ICC is an analysis of variance calculation in which the dependent variable is ratings and the independent variables are reviewers and papers. ICC has different forms (Yaffee, 2003), depending on factors such as whether reviewers are considered fixed or random, or whether one is concerned with consistency or exact agreement. For this analysis, Case 3 for k reviewers, or ICC (3, k) was chosen because reviewers analyzed are the only judges of interest. In Case 3, each target (paper) is evaluated by k (3) raters. In this study, the judges are a fixed effect, while the paper ratings are a random effect—a configuration known as the two-way mixed model. Brannick (2007) points out that ICC (3,k) is essentially equal to Cronbach’s alpha. Yet, as Shrout and Fleiss (1979) show, for each case of ICC there are two versions of reliability: (a) single measures, where the unit of analysis is the individual rating, and (b) average measures, where the unit of analysis is the mean of all the ratings. To find the second version of reliability, a Spearman-Brown correction is applied.

ICC ANOVAs were calculated to establish inter-rater reliability for each of the three dimensions evaluated by experts: (a) flow, (b) logic and (c) insight. Results, shown in Table 2, indicate high levels of inter-rater reliability for each dimension.
Table 2
*IntraClass Correlation Coefficients (ICC) for Expert Scores*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Average measures ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow</td>
<td>28.08</td>
<td>(116, 232)</td>
<td>.00</td>
<td>.766</td>
</tr>
<tr>
<td>Logic</td>
<td>13.29</td>
<td>(116, 232)</td>
<td>.00</td>
<td>.776</td>
</tr>
<tr>
<td>Insight</td>
<td>18.16</td>
<td>(116, 232)</td>
<td>.00</td>
<td>.794</td>
</tr>
</tbody>
</table>

Procedures

The assignment for participants was to write two drafts of a 5-6 page academic position paper on an educational psychology learning theory and to submit these drafts to the SWoRD System. In addition, participants were asked to review 5-6 peer submissions of each draft. For detailed description, see assignment sheet in Appendix B. While students received writing and reviewing scores for each draft, only writing scores on the first draft are relevant to the questions of this study.

The 174 students enrolled in Educational Psychology for the Spring 2005 semester were randomly assigned into six groups using a random number table. Even though the course is naturally divided into ten sections, the researcher circumvented these naturally-formed groupings in favor of randomly assigning treatments across the entire pool of 174 students. Therefore, students within each section received instructional packets which placed them into one of the six treatment groups. The nature of the assignment offers the necessary flexibility to have students within each section of the course working on different processes at different times.
To minimize the potential for teacher effect and to ensure consistency of procedures, the researcher introduced the assignment to each of the ten sections of the class. During the 45-minute introduction, instructional packets were distributed to participants in manila envelopes. Each envelope included the assignment sheet, (see Appendix B), a list of possible topics (Appendix C), the specific instructional packet for a given treatment (see Appendices D-I), a consent letter (Appendix J), and a printed copy of the SWoRD Version 3.0 Student Manual, available electronically at http://sword.lrdc.pitt.edu/getstarted.aspx. To follow proper procedure for working with human subjects in research, participants were informed of the nature of the study and were asked to sign the consent form (see Appendix J).

*Treatments*

One of the goals of this investigation is to measure differences in writing quality between participants who engage in reviewing activities prior to writing and those who write without having reviewed peer papers. Because the review-prior-to-writing condition does not occur naturally in the standard SWoRD cycle, approximately one third of the 174 students were assigned to write drafts earlier than the other groups. The first drafts generated by this group (Group One) provided review material for the entire group (n=57) of review-first treatments. To avoid a time-of-semester confound, none of these students were part of the actual subject pool for this investigation. Therefore of the six logistical groups, only five provided data for this analysis. The remaining 114 students were divided equally between a review-first (n=55) condition and a write-first condition (n=59). This grouping forms the basis of Phase I of the study. Phase II concerns only the
review-first group. The 2x2 factorial distribution of examples and feedback elaboration treatments was nested within the review-first group, breaking this larger group into three groups of n=14 and one group of n=13. See Appendices K and L for tabular and graphical representations of these schedules and distributions. The following paragraphs explain the five treatments of the investigation.

*Write First*

Group Six (n=59) represents the appended control group, designed to test for effects of reviewing prior to writing the first draft. This group followed the standard SWoRD sequence of steps: (a) write and submit draft one, (b) prepare for reviewing by viewing examples of helpful and unhelpful feedback and complete accompanying questionnaire (see http://peerfeedback.net and Appendix M), (c) review peers’ draft one, including rating student papers on a seven-point Likert scale as well as providing written comments on three dimensions of writing: flow, logic, and insight, (d) write and submit draft two, (e) back-review reviewers from first draft, (f) review peers’ draft two, and (g) back-review for draft two (see Appendix H). However, in order to isolate the effects of reviewing, this investigation only considers draft one activities, therefore the rest of the group sequences will not include steps after the submission of draft one. Groups 2-5, described below, share the common trait of having the order of writing and reviewing reversed to place reviewing papers before writing the first draft. Combined, these groups have n=55.
Review First, With Examples, With Elaboration

Group Five (n=14) followed all of the steps of the appended control group, with the exception of reversing the order of writing and reviewing. The sequence proceeded in the following order: (a) prepare for reviewing by viewing examples of helpful and unhelpful feedback and complete accompanying questionnaire, (b) review peers’ draft one, including rating student papers on a seven-point Likert scale as well as providing written comments (c) write and submit draft one, (d) write and submit draft two, (e) back-review reviewers from first draft, (f) review peers’ draft two, and (g) back-review for draft two (see Appendix G).

Review First, With Examples, Without Elaboration

Group Four (n=14) followed all of the steps of the Group Five, with the exception of viewing examples of feedback. Their sequence proceeded in the following order: (a) review peers’ draft one, including rating student papers on a seven-point Likert scale as well as providing written comments (b) write and submit draft one (see Appendix F).

Review First, Without Examples, With Elaboration

Group Three (n=13) did not engage in the examples activity but did provide the standard degree of feedback elaboration. Their sequence proceeded in the following order: (a) prepare for reviewing by viewing examples of helpful and unhelpful feedback and complete accompanying questionnaire, (b) review peers’ draft one, rating student papers on a seven-point Likert scale only, without comments (c) write and submit draft one (see Appendix E).
Review First, Without Examples, Without Elaboration

Group Two (n=14) did not engage in the examples activity and also did not provide the standard degree of feedback elaboration. Their sequence proceeded in the following order: (a) review peers’ draft one, rating student papers on a seven-point Likert scale only, without comments (b) write and submit draft one (see Appendix D).

Experimental Design and Data Analyses

Research Hypotheses

As addressed in Chapter I, this study seeks to examine the following hypotheses:

Phase I

1. Articulation of feedback by reviewers using a web-based peer review system for writing will result in higher quality of reviewers’ subsequent writing.

Phase II

1. The provision of examples of helpful and unhelpful reviews to reviewers will result in higher quality of reviewers’ subsequent writing.

2. Providing elaborate comments in addition to reviewer ratings will result in higher quality of reviewers’ subsequent writing.

3. There will be an interaction between example exposure type and feedback elaboration type.

Design

The hypotheses were tested in two phases. Phase I is a single factor design, where the independent variable is reviewing experience prior to writing with the following two levels: (a) reviewing prior to writing and (b) no reviewing prior to writing. Phase II is a
completely crossed 2 x 2 factorial design, where the independent variables are: (a) use of examples to guide student review (included versus excluded) and (b) level of feedback elaboration required by reviewers (elaborate versus simple). All four treatment groups reviewed peer papers prior to writing first drafts of their own.

The dependent variable is a composite writing score derived from combining three separate writing dimension scores. Dimension scores were averaged Likert scores on a seven-point scale in the following dimensions of writing: (a) flow, (b) logic, and (c) insight. These scores were added and converted to a total writing score. The dependent variable was measured on first draft only and represents the average of three expert scores. While both composite and individual scores were available for analysis, composite scores were deemed most appropriate due to the high correlations of the separate dimension scores. Flow and logic were correlated at .861; flow and insight were correlated at .889; and logic and insight were correlated at .880.

Though correlation is often the justification for MANOVA (Weinfurt, 2000), there are very few guidelines as to a range of correlation appropriate for this type of analysis. Maxwell (2001) asserts that a range of about .3 to about .7 of pairwise correlation is a good subjective rule of thumb, because anything lower suggests that variables are not related and anything higher is redundant. For highly correlated dependent variables, some argue that collapsing separate scores into a composite score and analyzing through univariate ANOVA is preferable (French, Poulsen, & Yu, 2002; Maxwell, 2001). Maxwell (2001) argues that if all the dependent variables are pairwise
correlated greater than .7, that it is preferable to create a composite score and execute a univariate ANOVA.

As discussed in the Writing Curricula and Research and Measuring Writing Quality sections of Chapter II, writing is a complex system of inter-related sub-processes that are often broken down into separate traits during analysis and assessment (Bellamy, 2006; Bereiter and Scardamalia, 1987). Yet, it is understood that the separate components of the writing process and its subsequent evaluation are often highly correlated. In fact, one of the most popular methods of evaluating writing over time has been holistic grading, which favors aggregation of all dimensions of writing due to their high levels of correlation.

Analyses

A priori, the analysis deemed most appropriate for this investigation was a multivariate analysis of variance or MANOVA (Weinfurt, 2000). There were several reasons MANOVA was an appropriate choice. First, existing literature suggests that particular dimensions of writing quality may be affected differently when a writer is exposed to models of peer writing (Charney & Carlson, 1995). Second, the primary justification for using MANOVA is correlation of dependent variables (Weinfurt, 2000, Stevens, 1996). Bray and Maxwell (1982) explain that MANOVA is most often used when a researcher is interested in evaluating “a set of measures as they represent some underlying construct” (p. 341). In this case, the set of measures is the set of writing dimension scores, and the underlying construct is writing quality.
Yet, during the course of the investigation, compelling reasons emerged for reconsidering this position. For example, dependent variables were predicted to be correlated, but only post-hoc analysis would reveal the high level of correlation among the dependent variables of this study ($r = .861-.889$). Therefore, following Maxwell’s (2001) recommendation, when correlations were found to be above .7 for all pairwise combinations of dependent variables, the researcher favored collapsing separate writing dimension scores into a composite score to be analyzed through univariate ANOVA. Also, Maxwell (2001) argues that typically when MANOVA is employed, it is mistakenly used to test hypotheses that are truly univariate. Since the questions investigated in the current context were univariate questions, ANOVA was deemed a more appropriate analysis. Given this rationale, the single factor design of Phase I was analyzed with a two-group ANOVA, and the 2 x 2 factorial design of Phase II was analyzed with a factorial ANOVA. However, for the sake of comparison, and because there is not complete consensus about the criteria for justifying one analysis over another, the data were also analyzed using MANOVA, the results of which are presented in Appendix P.

Prior to each test, the researcher tested for the assumptions of ANOVA, specifically, normality and homogeneity of variance. For the two-group ANOVA of Phase I, a one-sample Komogorov-Smirnov test revealed that data had adequate normality, $Z=1.02$, $p=.11$. For the factorial ANOVA of Phase II, a one-sample Komogorov-Smirnov test revealed that data had adequate normality, $Z=1.13$, $p=.15$. Levene’s test of equality of error variances reveal that homogeneity of variance is not
tenable $F(3, 51) = 5.10, p = .00$, however, the omnibus F-test is thought to be robust against heterogeneity of variances, especially with fixed effects and equal sample sizes (Box, 1954; Glass, Peckham, & Sanders, 1972; Hair, Anderson, & Black, 1998). Still, Moore (1995) suggests checking the ratio of variance between the smallest value and the largest value. A value of 4:1 or less is used as an informal standard. The ratio in this case is 3.73:1, so the ANOVA results are considered robust for this study. After ANOVA tests, since Phase I includes a single factor, and Phase II includes two factors, no post hoc analyses were necessary.
CHAPTER IV

RESULTS

Overview

Results will be presented for Phase I then Phase II of the study. Phase I was concerned with whether the articulation of feedback in a web-based peer review system leads to differences in the quality of subsequent writing by reviewers. Phase II of the study was concerned with the effects of varying aspects of reviewing activities, in particular levels of elaboration and levels of exposure to prototypal review examples. For each phase, ANOVA results for average expert scores are presented.

**Phase I – One Factor ANOVA**

The research hypothesis for Phase I was that articulation of feedback by reviewers using a web-based peer review system for writing would result in higher quality of reviewers’ subsequent writing. The hypothesis was analyzed through a single factor ANOVA with writing scores as a dependent variable. ANOVA assumptions were tenable. Table 3 presents the means and standard deviations for writing scores by treatment group. Means differ by only 1.41 on a 100 point scale (review-first, \(M = 47.04, \ SD = 19.32\); control, \(M = 48.45, \ SD = 18.90\)). ANOVA results indicate no significant differences for writing scores based on this treatment: \(F(1,112) = .16, p = .69\). It is worth noting that members of the control group outscored members of the treatment group, though not at a statistically significant level.
Table 3
Means and Standard Deviations of Writing Scores by Treatment

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>Sd</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review-First</td>
<td>47.04</td>
<td>19.32</td>
<td>55</td>
</tr>
<tr>
<td>Write-First</td>
<td>48.45</td>
<td>18.90</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td>47.77</td>
<td>19.03</td>
<td>114</td>
</tr>
</tbody>
</table>

Phase II – Factorial ANOVA

Preliminary Analyses

Preliminary analyses revealed that a few students had markedly low scores on the assignment. Upon examination of student artifacts for these students, it was discovered that three students had submitted assignments too incomplete to be evaluated by the dimensional criteria. For example, one student submitted a single-page bulleted list, when the assignment was to write a 5-6 page paper. Based on this discovery, three students were dropped from the analysis.

A one-sample Komogorov-Smirnov test revealed that data had adequate normality, \( Z=1.13, p=.15 \). Levene’s test of equality of error variances reveal that homogeneity of variance was not tenable, \( F (3,51) = 5.10, p = .00 \). As noted in Chapter III, Pearson product-moment correlations were calculated for the three writing dimensions, revealing a high degree of correlation (\( r = .861-.889 \)). Because these
correlations all fall above .7, ANOVA of composite writing scores was favored over MANOVA of separate writing dimensions (Maxwell, 2001).

After preliminary analyses, the following research hypotheses were tested in Phase II:

1. The provision of examples of helpful and unhelpful reviews to reviewers using a web-based peer review system for writing will result in higher quality of reviewers’ subsequent writing.

2. Providing elaborate comments in addition to reviewer ratings will result in higher quality of reviewers’ subsequent writing.

3. There will be an interaction between example exposure type and feedback elaboration type.

These hypotheses were analyzed through a 2 x 2 factorial ANOVA, where the independent variables were: (a) use of examples to guide student review (with-examples versus without-examples) and (b) level of feedback elaboration required by reviewers (with-elaboration versus without-elaboration), and the dependent variable was a composite writing score.

Hypothesis 1

Hypothesis 1 predicted that mean writing scores in the with-examples treatment group would be higher than those in the without-examples group. As shown in Table 4, mean differences were in the opposite direction from those predicted (with-examples, $M = 43.71$, $SD = 17.72$; without-examples, $M = 50.50$, $SD = 20.61$). As shown in Table 5, these differences were not statistically significant: $F(1,55) = 2.18$, $p = .14$. These results
do not support the hypothesis that providing examples of prototypical helpful and unhelpful reviews would lead to higher quality of reviewers’ subsequent writing.

*Hypothesis 2*

Hypothesis 2 predicted that mean writing scores in the with-elaboration treatment group would be higher than those in the without-elaboration group. As shown in Table 4, mean differences were consistent with the predicted direction (with-elaboration, \(M = 54.79, SD = 21.78\); without-elaboration, \(M = 39.57, SD = 13.09\)). As shown in Table 5, these differences were statistically significant: \(F(1,55) = 10.21, p = .00\). Partial Eta squared was .17. These results support the hypothesis that providing elaborate comments in addition to simple numeric ratings would lead to higher quality of reviewers’ subsequent writing.

*Hypothesis 3*

Hypothesis 3 predicted an interaction between examples and elaboration treatments. As shown in Table 4, means are ordered consistently across treatment groups. Each without-examples group mean is higher than its respective with-examples mean,  and each with-elaboration group mean is higher than its respective without-elaboration group mean (without-examples, with-elaboration \(M = 59.34, SD = 24.21\); without-examples, without-elaboration, \(M = 42.29, SD = 12.56\); with-examples, with-elaboration \(M = 50.57, SD = 19.19\); with-examples, without-elaboration, \(M = 36.85, SD = 13.50\)). As shown in Table 5, these differences were not statistically significant: \(F(1,55) = 38.02, p = .12\). These results do not support the hypothesis that there would be an interaction between examples and elaboration treatments.
### Table 4

**Means and Standard Deviations by Factorial Cell**

<table>
<thead>
<tr>
<th></th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Only</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating Only</td>
<td>$M = 36.85$</td>
<td>$M = 42.29$</td>
<td>$M = 39.57$</td>
</tr>
<tr>
<td></td>
<td>$SD = 13.50$</td>
<td>$SD = 12.56$</td>
<td>$SD = 13.09$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 14$</td>
<td>$n = 28$</td>
</tr>
<tr>
<td><strong>Rating and Commenting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating and Commenting</td>
<td>$M = 50.57$</td>
<td>$M = 59.34$</td>
<td>$M = 54.79$</td>
</tr>
<tr>
<td></td>
<td>$SD = 19.19$</td>
<td>$SD = 24.21$</td>
<td>$SD = 21.78$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 13$</td>
<td>$n = 27$</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M = 43.71$</td>
<td>$M = 50.50$</td>
<td>$M = 47.04$</td>
</tr>
<tr>
<td></td>
<td>$SD = 17.72$</td>
<td>$SD = 20.61$</td>
<td>$SD = 19.32$</td>
</tr>
<tr>
<td></td>
<td>$n = 28$</td>
<td>$n = 27$</td>
<td>$n = 55$</td>
</tr>
</tbody>
</table>

### Table 5

**Factorial ANOVA - Main Effects and Interactions for Examples and Elaboration**

<table>
<thead>
<tr>
<th></th>
<th>$MS$</th>
<th>$df$</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examples</strong></td>
<td>692.61</td>
<td>1</td>
<td>2.18</td>
<td>.15</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Elaboration</strong></td>
<td>3251.65</td>
<td>1</td>
<td>10.21</td>
<td>.00</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>38.02</td>
<td>1</td>
<td>.12</td>
<td>.73</td>
<td>.00</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>318.39</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

These results are not supportive of the hypothesis of Phase I: that articulation of feedback by reviewers using a web-based peer review system for writing would result in higher quality of reviewers’ subsequent writing. For Phase II, results were not supportive of hypothesis 1—that exposure to prototypical examples of feedback during reviewing would result in better subsequent writing. Hypothesis 2—that students who provided more elaborate feedback (ratings and comments) would perform better on the writing assignment than students who provided ratings alone—was supported. Finally, hypothesis 3, that there would be an interaction between examples and elaboration treatments, was unsupported. As noted in Chapter III, the data were also analyzed using MANOVA, the results of which are presented in Appendix P.
CHAPTER 5

DISCUSSION

This chapter will consider implications of results of the study. The chapter begins with an overview of the research questions and findings, then proceeds through the following topics: (a) discussion of the role of context within the current study, especially with regard to scoring, (b) consideration of the findings in light of existing research, (c) explanations of failed findings, (d) implications of study on pedagogical practice, (e) limitations of study, and (f) recommendations for further research.

Phase I of the study was concerned with the following research question:

1. Does the articulation of feedback in a web-based peer review system result in higher quality of reviewers’ subsequent writing?

ANOVA results indicated no differences between the review-first group and the write-first group.

Phase II was concerned with the impact of varying certain aspects of review activities. In this phase, the following questions were examined:

1. Does the provision of examples of helpful and unhelpful reviews result in higher quality of reviewers’ subsequent writing?

2. Does more elaborate feedback by reviewers result in higher quality of reviewers’ subsequent writing?

3. Is there an interaction between example exposure type and feedback elaboration type?
ANOVA results indicated no significant differences between groups attributable to examples treatments and also no interaction between examples and elaboration treatments. ANOVA results indicated main effects for elaboration treatments, supporting the hypotheses of question 2.

The Role of Context

In many ways, the SWoRD system is a departure from conventional methods of managing student writing assignments. The system’s most notable differences are that it is facilitated entirely online and that students’ grades for an assignment are determined by their peers. Anecdotal evidence suggests that students generally did not have too much trouble with the technical requirements and logistics of the system. For example, during the eleven weeks the assignment was active, there were relatively few questions related to website access, creation of online identification, or the submission of electronic files. This observation is perhaps not surprising given that the majority of participants were 1-2 years out of high school, likely placing them squarely in the category of digital natives (Prensky, 2001).

Yet, while students may be accustomed to using a computer to access the Internet and create word processing documents, they may not be accustomed to a computer-mediated system handling nearly every aspect of an assignment, including grading. Wasley (2006) writes of serious contention among faculty and students required to use a similar system in the freshman composition at Texas Tech University. Also, the novelty of an instructional medium has been shown to affect students’ ability to store information in memory. In his study of the instructional use of television, Singer (1980) noted
“pressures of attention to novel stimulation can actually interfere with our ability to store [information],” (p. 38). In the current context, the novelty of this system, combined with students’ awareness that the assignment was connected to a research project, may have had implications for student attitudes and engagement.

Because SWoRD activities do not happen during class time, there is potential that some students might feel less connected to a support structure, and therefore more taxed by the assignment as they independently make use of instructions and other resources to navigate the requirements of the assignment. If students were feeling disconnected or unsupported, the context of this study may have exacerbated such feelings. In order for the researcher to randomly assign treatments, instructions differed from student to student within sections of the class. Therefore, students’ ability to rely on classmates for help might have been severely limited for this assignment. Further, because treatments altered some of the standard steps of the typical SWoRD process, students had to pay extra close attention to their supplemental instructional packets, which differed from online instructions. For example, online instructions may have asked students to rate and comment on peer papers, but printed materials may have directed students to ignore the online direction to comment.

Though printed supplemental instructions were condensed into a simplified checklist (see Appendices D-I), anecdotal evidence suggests that students struggled to follow directions. For example, based on treatment group, students were instructed to sign up for a particular combination of writing assignment and reviewing assignment, which were labeled red and blue. Despite the simplified checklist and treatment-specific
reminder emails, around 30 students of 180 either made the wrong choice or had to email questions to make the correct choices. As students signed up for accounts and selected assignment options, instructors were able to monitor selections and have incorrect choices adjusted by the SWoRD system administrator. Still, these observations suggest that it may have been taxing on students to keep track of the differences between standard online instructions and their treatment-specific set of printed instructions.

Another aspect of the assignment that may have required adjustment on the part of participants is related to assignment grading. Aside from being put into the uncommon position of evaluating peer work, students were made aware that their own grades for this assignment would be determined entirely by their peers’ assessment of their work. Given the validity and reliability of SWoRD scores as found by Cho, Wilson, and Schunn (2006), objective consumers of such data are likely to put faith in the fact that valid and reliable grades can be calculated from the averaging of novice observations. Readers of Surowiecki’s popular book *The Wisdom of Crowds* (2004) may recognize a familiar theme—that the aggregation of novice observations often lead to elevated levels of accuracy in observations relative to the observations of single experts. However, students whose grade for a course is determined at around 30% from such a system are not likely to be as objective as non-participants. Indeed, student protests of the assignment generally centered on the issue of grading. Ultimately one student would file a formal complaint about her grade.

Since objections to grading methods were predicted by the researcher, information about the Cho, Wilson, and Schunn (2006) study, as well as information
about how grades are calculated in SWoRD, were explicitly discussed during the introduction to the assignment (see Appendix B). For example, it was explained to students that 50% of their review scores would be based on their accuracy, which would be calculated by how close they were to the mean review score. This description, while true, may have inadvertently encouraged students to make fewer distinctions between papers, instead favoring central positions on the Likert scale in an attempt to avoid rating too far from the mean. The data supports this hypothesis. As shown in Figure 1 and Figure 2, SWoRD (student) scores were grouped much more tightly around the mean than expert scores.

*Figure 1. Distribution of SWoRD Scores. $M = 79.71, SD = 6.82, n = 55.$*
Figure 2. Distribution of expert scores. $M = 47.04$, $SD = 19.32$, $n = 55$.

Information about how SWoRD calculates grades is readily available in online information about the system and particularly in the student and teacher instruction manuals. Therefore, it is possible that general users of SWoRD might respond, as described above, by not venturing too far from the middle of the Likert scale during reviewing. However, the context of the current study may have encouraged this response more than usual. Specifically, because the assignment was a requirement of the class and at the same time attached this research project, a fact known to all participants, extra effort was extended during the introduction of the assignment to address potential resistance toward the grading mechanisms. Also, for the sake of consistency while introducing the assignment, the researcher presented the materials to participants, rather than have them presented by the instructors for each section of the course. Having an
external figure introduce and explain an assignment that will contribute significantly
toward real student grades was one more reason for students to be suspicious of the
process or resistant to the assignment.

Since these potential reactions were considered prior to administering the
assignment, the nuances of the SWoRD system were very carefully explained in an
attempt to make students more comfortable about engaging in this unconventional system
for managing an assignment. In other words, the novelty of the system was reason
enough to predict student resistance and address it, but when coupled with a research
project, there were even more reasons to predict potential resistance. Therefore, during
the introduction of the assignment the extra attention paid to system logistics may have
drawn students into a strategic stance with regard to the system. Thus, hearing that
student review scores were based 50% on accuracy relative to the mean (in this case an
unknown quantity to them) might have been a signal to keep ratings toward the middle of
the scale. Also, students may have engaged in a tacit agreement to inflate ratings in order
to “beat the system.”

These hypothetical explanations could not be tested in the scope of this study, but
because the skew and kurtosis of the data could support these hypotheses, the researcher
opted to use expert scores for analyses rather than SWoRD scores. If the assignment had
not counted toward student grades, students may have had questionable motivation to
take it seriously, however, because it did count for a significant portion of students’
course grade and because the research context of the assignment may have influenced
student motivations, SWoRD scores in this instance may have been less valid and reliable
than in other contexts. Further, time did not permit the use of an optional calibration step that would have given students practice reviewing and feedback on the quality of their reviews prior to them counting. Applied research often must be accomplished within the limitations of the setting or context. In this case, expert scores provided a means for controlling for context effects. Analyses of SWoRD scores are available in Appendix O.

Expected Results

The results of the study indicate the strongest effects were related to the elaboration treatment. Those familiar with the SWoRD system know that reviews include both numerical Likert ratings and comments. Elaboration groups were simply asked to follow this standard format, which includes commenting. Rating-only groups were asked only to rate on the Likert scale and to type the word “blank” into the comments field. The variation allowed effects of providing comments to be separated from the effects of mere exposure to peer writing, even when it includes evaluative rating. It was hypothesized that the articulation of comments on specific dimensions of writing during peer review would lead to stronger cognitive organization for reviewers which, in turn, would translate to stronger writing of their own. Results of the study support the hypothesis. Members of the elaboration groups performed better than members of the without-elaboration groups in each of the evaluated dimensions of writing. These results support the theoretical positions of Britton (1970), Zinsser (1988) and others regarding the strong correlation between articulation and thought. These results also expand explanation literature, which has tended toward examining effects of self-explanation, most often in science and math. In developing comments for peers, it seems, students are
forced to organize their evaluative facilities enough to improve their own subsequent writing. In the current context, 17% of the variability was attributable to elaboration.

Unexpected Results

Review-first Versus Write-first

The two-group analysis of Phase I, comparing the writing performance of students who reviewed prior to writing with performance of those who did not review prior to writing, revealed that no differences could be attributed to the act of reviewing. At first glance, these results may seem to contradict support for the articulation hypotheses cited above. In other words, if reviewing is largely an act of articulating feedback, we would expect that students who reviewed papers prior to writing would perform better than those who did not. However, a closer look at the design of the study as well as the results from the factorial Phase II, provides some plausible explanations for the discrepancies.

First, it is important to consider that the composition of the review-first group includes each of the four groups from the 2 x 2 factorial analysis of Phase II. Recall the following treatment groups from the review-first condition: (a) with-examples and with-elaboration, (b) with-examples and without-elaboration, (c) without-examples and with-elaboration, and (d) without-examples and without-elaboration. Of these four groups, only the first group (with-examples and with-elaboration) follows the standard set of SWoRD reviewing activities that include exposure to prototypical examples of feedback as well as providing ratings and comments on peer papers. Each of the other three factorial groups nested in the larger review-first group is subtractive in at least one aspect from this first group. These three groups either forgo examples, elaboration, or both.
Therefore of the 55 participants in the review-first group, only 14 engaged in the full set of standard review activities.

Given the results of the factorial analyses, it is clear that variations in review treatments have an influence on the effects of reviewing. For example, since elaboration treatments in this study showed positive effects on reviewing, the 28 members of the review-first group who did not provide comments likely lost at least some of the benefits associated with reviewing. This effect may have been compounded by the fact that examples treatments in this context were shown to have negative effects on writing, though not at a significant level. Thus, only the without examples/with elaboration group (n=13) would have seen obvious advantages from the reviewing activities. Some of the other nested groups may have even acquired disadvantages. The fact that without-elaboration groups outperformed without-elaboration groups suggested that mere exposure to peer writing was much less powerful than when it involved the articulation of comments. These findings illuminate flaws in the design of the two-group analysis of Phase I, specifically that mere exposure to peer papers was the only common bond among the nested groups. In light of the data, it is not so surprising that the review-first group (n=55) as a whole did not outperform the write-first group (n=59). Future researchers in this area may favor a clean two-group comparison over nested treatments.

Examples

The most puzzling results from this study involve the examples treatment. Results of the factorial ANOVA of Phase II indicate that the members of the without-examples group performed better than members of the with-examples, though not at a significant
level ($p = .15$). Recall that with regard to examples, members of the review-first group were placed in either a with-examples group or a without-examples group. The with-examples group was instructed to visit the following website: http://peerfeedback.net where they would watch a multimedia presentation about how to give good feedback on writing. The multimedia presentation is an eleven-slide, eleven-minute, narrated slide show presenting, through text and voiced-over characters, seven bad cases of writing and two good cases. A narrator sets up each case before characters read quoted examples of reviewer feedback and corresponding author responses. Each slide contains only text, and many are quite lengthy, the longest being nearly 300 words. Students were asked to complete a reaction sheet after viewing the presentation. The reaction sheet was considered sufficient proof that students followed the treatment. While the multimedia presentation is typically recommended on the SWoRD site, administrators of the site agreed to remove traces of this recommendation during the period of this study. Thus, only members of the with-examples group were given the link to http://peerfeedback.net.

There is little in the content of the slide presentation that seems likely to explain lower subsequent writing scores. Table 6 shows content of feedback examples as well as their didactic purpose. The first column presents the target of criticism found in example review comments. The second column presents the problem with the specific comments made by example reviewers. So, the second column is in essence a record of the purpose of a given example. Because the goal of the multimedia presentation is to illuminate problems with the form of comments made, the targets of criticism in these sample comments seem to be immaterial. For example, if the purpose of an example is to
illustrate the need for specificity in comment-giving, it matters little whether the example
happens to be about grammatical errors, structure problems, or diction. However, if
viewers of the presentation were absorbing the arbitrary target of sample criticism, it is
possible that they may have been led to focus more on superficial aspects of the writing,
because that is where most of the sample critiques were pointed. It is also possible that
viewers made mental notes about being concise and avoiding extraneous information,
since these were problems noted in a few of the sample critiques (see Table 6). While
there is no way to tell if members of with-examples groups were influenced by the
content of the examples, it is a possibility worth considering.

<table>
<thead>
<tr>
<th>Target of feedback criticism</th>
<th>Problem with Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical errors</td>
<td>Not specific enough</td>
</tr>
<tr>
<td>Extraneous information</td>
<td>Not specific enough</td>
</tr>
<tr>
<td>Points need made clearer</td>
<td>No justification</td>
</tr>
<tr>
<td>Poor rating, but no problems noted</td>
<td>No justification</td>
</tr>
<tr>
<td>No new insights</td>
<td>Insulting comments</td>
</tr>
<tr>
<td>Sources not cited</td>
<td>Insulting comments</td>
</tr>
<tr>
<td>Good rating and comments</td>
<td>Nothing to improve</td>
</tr>
<tr>
<td>Specifics needed, diction not concise, structure</td>
<td>Many specific, respectful suggestions</td>
</tr>
<tr>
<td>problems</td>
<td></td>
</tr>
<tr>
<td>Too much extraneous material</td>
<td>Many specific, respectful suggestions</td>
</tr>
</tbody>
</table>
Another possible explanation for the results of the examples treatments is that examples treatments made reviewing too easy. The multimedia presentation of feedback examples was quite detailed, and may have supplied members of this group with clear templates for providing feedback. On one hand, such templates may have facilitated review processes for students who, as has been noted, have little experience evaluating peer work. On the other hand, if these templates were employed, they may have inhibited the need for participants to engage in the process of generating original articulation of evaluative observations.

A different perhaps more plausible explanation for results of the with-examples group is that the activity was a burdensome additional requirement embedded into a very taxing assignment. It is possible that many students were nearing a mental effort threshold with the larger SWoRD assignment when this additional activity was introduced. Up to this point, students had been asked to process and accept the following: (a) an unusual approach to managing an assignment, (b) controversial grading mechanisms which are supported by complex statistics, (c) two sets (one electronic and one printed) of very detailed and sometimes disparate sets of instructions, (d) a multi-page student manual about how SWoRD works, (e) the responsibility of accurately evaluating the work of one’s peers, (f) implications for involvement in a research study, (g) interactions with a support structure and support personnel external to the class, and (h) the usual requirements of a writing assignment, such as reading research, planning, drafting. It is possible that with-examples group members may have reached limits as to
how much energy they could dedicate toward this one assignment in this one particular class.

Salomon (1983) described mental effort as controlled, non-automatic elaborations applied to materials being learned. Since such effort involves choice, student motivation has a causal relationship with mental effort—a relationship sometimes informed by attributions and expectations for success in learning (Cennamo, 1989). According to Weiner (1979), students will persist in a task if they consider success or failure to be related to unstable causes, such as effort or luck, as opposed to stable causes, such as task difficulty. In other words, if a task is perceived to be so difficult that students become convinced that even effort or luck will not influence outcomes, they will not likely persist. Weiner (1979) suggests that students work best at tasks of intermediate difficulty.

One might find clues as to the merit of mental effort conjecture in the review comments produced by students from the examples group. For example noticeable differences in the length and complexity of comments generated by group members might support an effort threshold hypothesis. A rudimentary analysis of the comments rendered by students is only mildly supportive of this hypothesis, however. The mean of total words generated per reviewer across with-examples and without-examples groups is 1184. On average, members of the with-examples group produced only 18 fewer words than their without-examples counterparts ($M=1176$, $M=1194$ respectively). Therefore, the with-examples numbers were lower as predicted, but not significantly. Note that these data do not provide definitive support for or against a mental effort threshold theory.
Implications of Study

Results of the study clearly support the use of explanation activities requiring students to articulate their evaluative observations of peer work. The with-elaboration groups out-performed without-elaboration groups on the writing assignment, and this treatment accounts for about 17% of the total variability. As noted in the literature on self-explaining (Chi et al., 1989; Chi & VanLehn, 1991; Ploetzner et al. 1999) teach to learn (Bargh & Schul, 1980; Cloward, 1967; Palinscar & Brown, 1984; Rekrut, 1992), and write to learn (Bereiter & Scardamalia, 1987; Zinsser, 1988), the act of articulating observations through language seems to organize conceptual information in a way that is accessible to students during later activities. In this case, articulations made during review activities led to differences in writing activities that were not completed until two weeks later. As noted by Leelawong et al. (2002), teachers’ knowledge structures are organized as much by answering student questions and providing feedback as by preparing materials to teach. When students engage in evaluative activities normally reserved exclusively for the teacher, they stand to see cognitive benefits. Articulation marks a commitment to an observation, which, as Kelly (1963) noted, allows one to predict, test and revise conceptions. While some researchers have argued for training students in peer review (Hu, 2005) or explanation strategies (Bielaczyc, Pirolli & Brown, 1995) prior to their engagement in these activities, results from the current study suggest there are some raw benefits associated with the activity of articulating comments, even in the absence of extensive training.
Limitations of Study

*Draft One as a Measure*

Since this study was primarily concerned with effects of reviewing, the scope of the study was constricted to a measure of first-draft writing. Focusing on first-draft writing prevented interference from other variables, such as time of semester, effects from writing a previous draft, and effects of back-reviews; however some students may have thought of the first draft as more of a *rough* draft. While multiple drafts are likely common for writing assignments, it might be less common to count each draft equally toward a final score. It is perhaps beyond the scope of this dissertation to include discussion of the merits of SWoRD’s equally weighted, two-draft design. Student motivation on the first draft was likely a consideration made by the designers. But, it has been mentioned here simply to illustrate that students’ responses to the concept of a graded first draft may have been influenced by past experience or expectation, despite explanations about how the first draft would be scored in the current context.

Random assignment likely distributed any first-draft/rough-draft interpretation effects equally throughout the groups, yet given the possibility that effort on the first draft may have been determined, to a certain extent, by one’s predisposition toward writing multiple drafts, second drafts may have been a truer measure of students’ writing skills. With draft two being the final version of the paper to be graded, students’ maximum effort for the assignment would have been accounted for by this point in time. However, measuring second-draft writing was prohibited by design, given the research questions, the goals of the study, and the requirements of the classes involved.
**Time**

Given that students are probably not typically put in the role of evaluator, it is logical that preparation and practice will lead to improved reviewing and writing. One way of preparing students for reviewing activities is through a calibration exercise. Typically calibration involves a practice round of reviewing on simulated papers, after which students receive feedback on their reviewing. Calibrated Peer Review (Robinson, 2001), an alternative to SWoRD, emphasizes calibration enough to include it in its name but also to weigh subsequent reviews based on calibration activity results. Thus, students who perform best in calibration will have their scores count more heavily in the real review scores that follow.

In SWoRD, calibration is an optional step recommended by the designers of the system. The step is thought to make students more comfortable about the reviewing process, but also to improve their review accuracy. Based on the amount of time already allocated to the SWoRD assignment, it was not possible include the calibration step in this assignment. Therefore, the potency of review activities for participants may have been reduced, possibly enough to at least partially explain the lack of differences between the review-first group and the write-first group in the two-group analysis of Phase I. Also, the omission of the calibration step presented one more reason to favor expert scores over student scores in this context.

**Maturity of Participants**

While the study was deliberately implemented in a core, required class for all undergraduate education majors so as to capture a representative sample of all
concentrations, the class is taken by sophomores, 96% of whom are below the age of 21. There is no way to know if results will generalize to other levels of education and other age groups. If there were issues related to mental effort, for example, it is possible that graduate students would have been affected differently by the complex directions. Motivation and experience likely play disparate roles for graduate students versus undergraduates.

**Diffusion of Treatments**

Internal validity was threatened by the lack of control for a diffusion of treatments effect. Usually this threat is due to the inability of a researcher to control the sharing of information between treatment groups. In this study, the threat is more likely to come from students inadvertently missing important directions. For example, there is no way to know if members of the without-examples groups did or did not view the multimedia presentation of examples of prototypal feedback. Even though it was not in either the short or long forms of their instructions (see Appendices D-H), it is possible they were pointed to the website by “helpful” classmates in the with-examples groups. Another, perhaps even more likely possibility for a diffusion effect, is that students in without-elaboration groups may have overlooked the direction in their particular checklist (see Appendices D and E) to leave comment fields blank, in which case they actually followed the steps prescribed for with-elaboration groups. However, because the SWoRD system stores these data, the researcher can accurately track if participants followed this particular treatment. Ten participants out of 27 in the without-elaboration group were found to have mistakenly articulated comments. When, for the sake of analysis, students
are moved into groups corresponding to what they actually did rather than what they were instructed to do, ANOVAs produce very similar results for elaboration treatments (see Appendix P), lending more support to the validity of reported results. Examples treatments could not be tracked with this degree of precision, and therefore could not be checked in the same way. Also, examples treatments did not involve competing sets of instructions, and so seem less vulnerable to diffusion of treatment effects.

Absence of Attitudinal Data

Since data regarding student perceptions and attitudes were not collected in this investigation, there is little data to support or reject speculation about mental effort. These data may have provided more information about whether students were consciously confused by directions or resistant to the overall assignment. An anonymous survey also may have revealed if students deliberately ignored treatment directions in order to acquire an advantage in performance. Such data may have helped inform interpretations of the quantitative analyses performed in the study.

Recommendations for Future Research

As per the discussion of the “role of context” (above), student attitudes toward and perceptions of grading mechanisms and other novel aspects of the SWoRD system, may have influenced their levels of motivation and engagement in the assignment. Survey data on student attitudes and perceptions may have provided more information about issues such as mental effort and how these factors may have affected results. Future researchers of SWoRD and other web-based reviewing systems may want to include this component for analysis.
However, ease of interpreting such self-reported data may depend on context. Previous measures of student attitudes and perceptions surrounding online peer review reveal that many students take issue with the medium, but their scores tell a different story (DiGiovanni & Nagaswami, 2001; Sullivan, Brown, & Nielson, 1998). MacLeod (1999) found that students preferred face-to-face peer reviews but were more candid on computer-mediated systems.

Given the unexpected results in the study, specifically the lack of differences between the review-first and write-first groups of Phase I, more research is needed to determine if these results carry beyond the context of this study. For example, a follow-up two-group study without factorial treatments nested in the review-first group might provide cleaner information about an overall effect of reviewing activities. Also, given the inability to control for diffusion effects in the examples treatment of the current study, it might be appropriate to perform future investigations of examples treatments in laboratory settings where one can exert more control over diffusion effects.
APPENDICES
APPENDIX A

SWoRD Review Form
BEFORE you review,
Please read your peer writing carefully with reference to the criteria given in this form. 50% of your reviewing grades are from your authors who will receive your feedback. She or he will decide how helpful your feedback would be in revising their writing.

Peer writing you are reviewing is …
Title: _________________________________________________
Author: _______________________________________________
Due by: _______________________________________________

WHILE you review,
There are two very important parts to giving good feedback. First, give very specific comments rather than vague comments: Point to exact page numbers and paragraphs that were problematic; give examples of general problems that you found; be clear about what exactly the problem was; explain why it was a problem, etc. Second, make your comments helpful. The goal is not to punish the writer for making mistakes. Instead your goal is to help the writer improve his or her paper. You should point out problems where they occur. But don’t stop there. Explain why they are problems and give some clear advice on how to fix the problems. Also keep your tone professional. No personal attacks. Everyone makes mistakes. Everyone can improve writing.

AFTER you review,
Note that this form DOES NOT automatically turn your reviews in. After you are done, you need to access to SWoRD. Then, please copy & paste your review results into the online form.

WHENEVER you review,
If you have any questions or comments, feel free to email Kwangsu Cho at kwangsu@pitt.edu.
1. Prose Flow
Did the writing flow smoothly so you could follow the main argument? This dimension is not about low level writing problems, like typos and simple grammar problems, unless those problems are so bad that it makes it hard to follow the argument. Instead this dimension is about whether you easily understood what each of the arguments was and the ordering of the points made sense to you. Can you find the main points? Are the transitions from one point to the next harsh, or do they transition naturally?

<table>
<thead>
<tr>
<th>Your Rating: Based on your comments above, how would you rate the prose flow of this paper?</th>
</tr>
</thead>
<tbody>
<tr>
<td>□  7. Excellent</td>
</tr>
<tr>
<td>□  6. Very good</td>
</tr>
<tr>
<td>□  5. Good</td>
</tr>
<tr>
<td>□  4. Average</td>
</tr>
<tr>
<td>□  3. Poor</td>
</tr>
<tr>
<td>□  2. Very poor</td>
</tr>
<tr>
<td>□  1. Disastrous</td>
</tr>
</tbody>
</table>

Your Comments: First summarize what you perceived as the main points being made so that the writer can see whether the readers can follow the paper’s arguments. Then make specific comments about what problems you had in understanding the arguments and following the flow across arguments. Be sure to give specific advice for how to fix the problems and praise-oriented advice for strength that made the writing good.
2. Logic Argument
This dimension is about the logic of the argument being made. Did the author just make some claims, or did the author provide some supporting arguments or evidence for those claims? Did the supporting arguments logically support the claims being made or where they irrelevant to the claim being made or contradictory to the claim being made? Did the author consider obvious counter-arguments, or were they ignored?

Your Comments: Provide specific comments about the logic of the author’s argument. If points were just made without support, describe which ones they were. If the support provided doesn’t make logical sense, explain what they is. If some obvious counter-argument was not considered, explain what that counter-argument is. The give a potential fixes to these problems if you can think of any. This might involve suggesting that the author change their argument.

Your Rating: Based on your comments above, how would you rate the logical arguments of this paper?

☐ 7. Excellent All arguments strongly supported and no logical flaws in the arguments
☐ 6. Very good All but one argument strongly supported or one relatively minor logical flaw in the argument.
☐ 5. Good All but two or three arguments strongly supported or a few minor logical flaws in arguments
☐ 4. Average Most arguments are well supported, but one or two points have major flaws in them or no support provided
☐ 3. Poor A little support presented for many arguments, or several major flaws in the arguments
☐ 2. Very poor Little support presented for most arguments, or obvious flaws in most arguments
☐ 1. Disastrous No support presented for any arguments, or obvious flaws in all arguments presented
3. Insight beyond core readings
This dimension concerns the extent to which new knowledge is introduced by a writer. Did the author just summarize what everybody in the class would already know from coming to class and doing the assigned readings, or did the author tell you something new?

Your Comments: First summarize what you think the main insights were of this paper. What did you learn if anything? Listing this clearly will give the author clear feedback about the main point of writing a paper: to teach the reader something. If you think the main points were all taken from the readings or represent what everyone in the class would already know, then explain where you think those points were taken or what points would be obvious to everyone. Remember that not all points in paper need to be novel, because some of the points need to be made just to support the main argument.

Your Rating: Based on your comments above, how would you rate the insights of this paper?

☐ 7. Excellent
   I really learned several new things about the topic area, and it changed my point of view about that area.

☐ 6. Very good
   I learned at least one new, important thing about the topic area.

☐ 5. Good
   I learned something new about the topic area that most people wouldn’t know, but I’m not sure it really important for that topic area.
   All the main points weren’t taken directly form the class readings, but most people but many people would have thought that on their own if they would have just taken a little time to think.
   Some of the main points were taken directly form the class readings; the others would be pretty obvious to most people in the class.

☐ 4. Average

☐ 3. Poor
   Most of the main points were taken directly form the class readings; the others would be pretty obvious to most people in the class.

☐ 2. Very poor
   All the points stolen directly from the class readings

☐ 1. Disastrous

Make sure that you have saved your reviews before you leave.
APPENDIX B

Assignment Description
Educational Psychology Theorist Research Paper

The Big Picture
For this assignment, you are asked to write two drafts of a 5-6 page (double-spaced) research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm.

Procedure Highlights
Your papers and your reviews must be submitted by the due date described in your individual packet and must be submitted electronically through an online peer review system called SWoRD (described below). You should submit your papers as Microsoft Word Documents or Adobe PDFs. (A packet of instructions for using the system is attached). Because this system relies on peer assessment, late papers cannot be accepted. You have a one-day grace period after the deadline of an assignment, during which time you can still submit your assignment with a 10% reduction in score. Papers submitted after the grace period will not be accepted. Students who miss first drafts cannot turn in final drafts. It will be important to turn in what you have on time even if it is not exactly where you want it to be. Please note that your reviews comprise a significant portion of your grade.

Description of the SWoRD Peer Review System
What is SWoRD? SWoRD stands for Scaffolded Writing and Rewriting in the Discipline. It is a web-based peer review system developed at the University of Pittsburgh that manages the submission and evaluation of student papers, including the distribution of papers to peer reviewers. The system’s evaluation instruments guide and calculate writing and reviewing scores.

A Contribution to the Field of Educational Psychology
This assignment is being connected to a research project for a doctoral dissertation. It is an important way to help make a contribution to the field, as this is how we learn about teaching and learning and how we progress as educators. You are not
obligated to participate in the study, however, since the study makes use of an existing assignment, you are required by your instructor to complete the assignment in exactly the same way, whether or not you agree to participate. Agreement to participate merely gives the researcher access to papers and scores for the purposes of measuring average scores. As per the attached consent form, participants are never identified in any subsequent presentation of results. The assignment will count toward your grade in the exact same way whether or not you participate in the study. There is absolutely no advantage or disadvantage to you based on this decision.

In this study, students are randomly assigned to different treatments, meaning that there are very minor variations in the assignment details depending on what treatment students are assigned. In research, random assignment ensures the validity of results. If we used existing groups, such as class sections, we can’t be sure if there are differences in these groups already or if differences are because of our treatments. For example, one might suggest that students enrolling in early classes are more eager and will tend to do better on assignments regardless of what treatments we use.

Again, we have used random assignment for this study. This means that instructions given to you may be slightly different than the person next to you, including details such as due date. For this reason, it will not help, and potentially could hurt if you ask peers in the class about assignment details. Further, discussing these details with peers can jeopardize the integrity of the research. Please do not discuss procedural questions with your classmates. Ask your instructor, the help desk, or the researcher (rwooley@kent.edu, 330-807-1718) if you have questions about procedures. If you lose your instructions or checklist, you must contact the researcher to get a new packet.

**Evaluation**

Your evaluation for this assignment, which includes two drafts of your paper as well as two sets of reviewing, is worth a total of 84 pts. This constitutes 33.6% of your grade in this course (except for Dr. Froehlich’s class). The points are broken down in the following table:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Dimensions of Evaluation</th>
<th>Method of Evaluation</th>
<th>Point Total for Assignment</th>
<th>Percentage of Course Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Drafts One and Two</td>
<td>Flow, Logic and Insight (each on a seven-point scale)</td>
<td>Peer evaluation</td>
<td>42 points Seven from each dimension - 21 points each between 1st and 2nd drafts (Basically double your SWoRD-Generated writing score for this number)</td>
<td>16.8%</td>
</tr>
<tr>
<td>Review Peer Drafts</td>
<td>Accuracy and/or</td>
<td>Accuracy calculation</td>
<td>32 points 16 points for each draft,</td>
<td>12.8%</td>
</tr>
</tbody>
</table>
Helpfulness and/or back review by peers that includes review accuracy and, if applicable, back review

| APA format | Accuracy | Instructor | 10 points | 4% |

Sometimes students question the reliability (consistency of the same grader) and validity (accuracy of evaluation) of peer assessments. These questions usually rest on a fallacy that instructor ratings are not prone to such problems. Research shows that the peer rating system used by SWoRD is even more reliable and valid than single instructor ratings, regardless of instructor expertise. Students tend to do a better job reviewing for a few reasons: 1. Because students have 5-6 papers instead of 25-200, they can focus on the papers better and tend to stay more consistent in their evaluations. 2. Because students are closer to the same cognitive level as their peers, they tend to give advice that is more understandable and useful to their peers. 3. Because 5-6 peers perform the reviews versus one faculty member, peer reviews tend to be more objective and accurate. (If this topic intrigues you, perhaps check out a book called *The Wisdom of Crowds* by James Surowiecki, which gives several examples of how groups of laypeople are almost invariably more accurate than single experts in everything from choosing winning stocks to choosing correct answers on *Who Wants to Be a Millionaire*.) Despite all of these advantages to peer review, there are additional safeguards in place to ensure that papers are evaluated correctly. First, peer ratings are weighted by their accuracy. Second, peers are accountable for their evaluations. And finally, all procedures are under close review by the instructors.
APPENDIX C

Topic Choices
Research Possibilities

BF Skinner, behavior learning theory
Paulo Freire, social justice
Marva Collins, educating inner city children
Albert Bandura, social learning theorist
John Dewey, progressive education
Barbara Rogoff, holistic approach to education
Robert Gagné, conditions of learning and instruction
Maria Montessori, child-centered education
Kenneth Gergen, social constructionist, student-led education
Jerome Bruner, social constructionist, Guided Discovery
Gallaudet, deaf education
Lev Vygotsky, social constructionist
Benjamin Bloom, taxonomy of educational objectives
H.D. Thoreau
Howard Gardner, multiple intelligences
Alfie Kohn
David Ausubel, advanced organizers
Elliot Eisner, art education
Urie Bronfenbrenner, social context of childhood
Jean Piaget, constructivist
Reggio Emilia, society supported, child-centered
John Anderson, memory
Other
Other
APPENDIX D

Group 2 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview
For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Assigned Peer Papers</td>
<td>2/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit First Draft</td>
<td>3/2/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:
After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley)Educational Psychology Kent State University”. When asked to select a topic, choose Topic 2 – Blue for writing and Topic 1 – Red for Reviewing. Be careful to select the right topics! Notice that you are selecting a different topic for writing than for reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist. *The checklist is very, very important.*
## Assignment Checklist

<table>
<thead>
<tr>
<th>Example</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>Read over assignment sheets and instructions</td>
</tr>
<tr>
<td>☐</td>
<td>Choose a topic from list and clear it with instructor</td>
</tr>
<tr>
<td>☐</td>
<td>Use SWoRD Student Manual and <a href="http://ladybug.lrdc.pitt.edu/sword3">http://ladybug.lrdc.pitt.edu/sword3</a> to register for the SWoRD system by <strong>February 10</strong>.</td>
</tr>
<tr>
<td>☐</td>
<td><strong>Note:</strong> When reviewing, rate papers with numeric scores only. Do not provide comments on papers you review. Leave comment fields blank. Your review score will be based only on your accuracy, so think about the numbers you assign.</td>
</tr>
<tr>
<td>☐</td>
<td>Review Assigned Peer Papers by <strong>February 16, 11:59 pm.</strong></td>
</tr>
<tr>
<td>☐</td>
<td>View the following website for plagiarism and APA guidelines: <a href="http://fpdc.kent.edu/PlagiarismAPA.htm">http://fpdc.kent.edu/PlagiarismAPA.htm</a></td>
</tr>
<tr>
<td>☐</td>
<td>Write first draft of paper, making sure to follow APA format</td>
</tr>
<tr>
<td>☐</td>
<td>Submit first draft of paper by <strong>March 2, 11:59 pm.</strong></td>
</tr>
<tr>
<td>☐</td>
<td>Check scores in SWoRD to gauge progress in class <strong>March 11</strong></td>
</tr>
<tr>
<td>☐</td>
<td>Submit Second Draft by <strong>March 16, 11:59 pm.</strong></td>
</tr>
</tbody>
</table>
| ☐ | Submit Back Reviews for First Draft by **March 23, 11:59 pm.**  
**Do not back review anyone with blank comment fields!** |
| ☐ | Review Assigned Peer Papers by **April 6, 11:59 pm.** |
| ☐ | Check scores in SWoRD to gauge progress in class **April 13** |
| ☐ | Submit Back Reviews for Second Draft by **April 13, 11:59 pm.**  
**Do not back review anyone with blank comment fields!** |
APPENDIX E

Group 3 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview
For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Assigned Peer Papers</td>
<td>2/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit First Draft</td>
<td>3/2/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:
After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley)Educational Psychology Kent State University”. When asked to select a topic, choose Topic 2 – Blue for writing and Topic 1 – Red for Reviewing. Be careful to select the right topics! Notice that you are selecting a different topic for writing than for reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist, including looking at sample files. *The checklist is very, very important.*
# Assignment Checklist

<table>
<thead>
<tr>
<th></th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>✅</td>
<td>Read over assignment sheets and instructions</td>
</tr>
<tr>
<td></td>
<td>Choose a topic from list and clear it with instructor</td>
</tr>
<tr>
<td></td>
<td>Use <em>SWoRD Student Manual</em> and <a href="http://ladybug.lrdc.pitt.edu/sword3">http://ladybug.lrdc.pitt.edu/sword3</a> to register for the SWoRD system by <strong>February 10</strong>.</td>
</tr>
<tr>
<td></td>
<td>Prepare for reviewing by viewing the following website for examples of reviews: <a href="http://www.peerfeedback.net">www.peerfeedback.net</a>. Complete accompanying question sheet labeled “Peerfeedback.net” and submit to instructor.</td>
</tr>
</tbody>
</table>
|   | Review Assigned Peer Papers by **February 16, 11:59 pm**.  
**Note:** When reviewing, rate papers with numeric scores only. Do not provide comments on papers you review. Leave comment fields blank.  
Your review score will be based only on your accuracy, so think about the numbers you assign. |
|   | View the following website for plagiarism and APA guidelines: [http://fpdc.kent.edu/PlagiarismAPA.htm](http://fpdc.kent.edu/PlagiarismAPA.htm) |
|   | Write first draft of paper, making sure to follow APA format |
|   | Submit first draft of paper by **March 2, 11:59 pm**. |
|   | Check scores in SWoRD to gauge progress in class **March 11** |
|   | Submit Second Draft by **March 16, 11:59 pm**. |
|   | Submit Back Reviews for First Draft by **March 23, 11:59 pm**.  
**Do not back review anyone with blank comment fields!** |
|   | Review Assigned Peer Papers by **April 6, 11:59 pm**. |
|   | Check scores in SWoRD to gauge progress in class **April 13** |
|   | Submit Back Reviews for Second Draft by **April 13, 11:59 pm**.  
**Do not back review anyone with blank comment fields!** |
APPENDIX F

Group 4 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview
For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm.

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Assigned Peer Papers</td>
<td>2/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit First Draft</td>
<td>3/2/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:
After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley) Educational Psychology Kent State University”. When asked to select a topic, choose Topic 2 – Blue for writing and Topic 1 – Red for Reviewing. Be careful to select the right topics! Notice that you are selecting a different topic for writing than for reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist. *The checklist is very, very important.*
### Assignment Checklist

<table>
<thead>
<tr>
<th>Task</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read over assignment sheets and instructions</td>
<td></td>
</tr>
<tr>
<td>Choose a topic from list and clear it with instructor</td>
<td></td>
</tr>
<tr>
<td>Use <strong>SWoRD Student Manual</strong> and <a href="http://ladybug.lrdc.pitt.edu/sword3">http://ladybug.lrdc.pitt.edu/sword3</a> to register for the SWoRD system by <strong>February 10</strong>.</td>
<td></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>February 16, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>View the following website for plagiarism and APA guidelines:</td>
<td><a href="http://fpdc.kent.edu/PlagiarismAPA.htm">http://fpdc.kent.edu/PlagiarismAPA.htm</a></td>
</tr>
<tr>
<td>Write first draft of paper, making sure to follow APA format</td>
<td></td>
</tr>
<tr>
<td>Submit first draft of paper by <strong>March 2, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>March 11</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Second Draft by <strong>March 16, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft by <strong>March 23, 11:59 pm.</strong></td>
<td><strong>Do not back review anyone with blank comment fields!</strong></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>April 6, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>April 13</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft by <strong>April 13, 11:59 pm.</strong></td>
<td><strong>Do not back review anyone with blank comment fields!</strong></td>
</tr>
</tbody>
</table>
APPENDIX G

Group 5 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview
For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm.

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Assigned Peer Papers</td>
<td>2/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit First Draft</td>
<td>3/2/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:
After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley)Educational Psychology Kent State University”. When asked to select a topic, choose Topic 2 – Blue for writing and Topic 1 – Red for Reviewing. Be careful to select the right topics! Notice that you are selecting a different topic for writing than for reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist, including viewing example files as assigned. The checklist is very, very important.
Assignment Checklist

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read over assignment sheets and instructions</td>
</tr>
<tr>
<td>Choose a topic from list and clear it with instructor</td>
</tr>
<tr>
<td>Use SWoRD Student Manual and <a href="http://ladybug.lrdc.pitt.edu/sword3">http://ladybug.lrdc.pitt.edu/sword3</a> to register for the SWoRD system by February 10.</td>
</tr>
<tr>
<td>Prepare for reviewing by viewing the following website for examples of reviews: <a href="http://www.peerfeedback.net">www.peerfeedback.net</a>. Complete accompanying question sheet labeled “Peerfeedback.net” and submit to instructor.</td>
</tr>
<tr>
<td>Review Assigned Peer Papers by February 16, 11:59 pm.</td>
</tr>
<tr>
<td>View the following website for plagiarism and APA guidelines: <a href="http://fpdc.kent.edu/PlagiarismAPA.htm">http://fpdc.kent.edu/PlagiarismAPA.htm</a></td>
</tr>
<tr>
<td>Write first draft of paper, making sure to follow APA format</td>
</tr>
<tr>
<td>Submit first draft of paper by March 2, 11:59 pm.</td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class March 11</td>
</tr>
<tr>
<td>Submit Second Draft by March 16, 11:59 pm.</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft by March 23, 11:59 pm. Do not back review anyone with blank comment fields!</td>
</tr>
<tr>
<td>Review Assigned Peer Papers by April 6, 11:59 pm.</td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class April 13</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft by April 13, 11:59 pm. Do not back review anyone with blank comment fields!</td>
</tr>
</tbody>
</table>
APPENDIX H

Group 6 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview

For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm.

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submit First Draft</td>
<td>3/2/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>3/9/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:

After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley) Educational Psychology Kent State University”. When asked to select a topic, choose Topic 2 – Blue for both writing and reviewing. Be careful to select the right topics! Notice that you are selecting the same topic for writing and reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist. The checklist is very, very important.
**Assignment Checklist**

<table>
<thead>
<tr>
<th>Task</th>
<th>Checkmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td></td>
</tr>
<tr>
<td>Read over assignment sheets and instructions</td>
<td></td>
</tr>
<tr>
<td>Choose a topic from list and clear it with instructor</td>
<td></td>
</tr>
<tr>
<td>Use <em>SWoRD Student Manual</em> and <a href="http://ladybug.lrde.pitt.edu/sword3">http://ladybug.lrde.pitt.edu/sword3</a> to register for the SWoRD system by <strong>February 24</strong>.</td>
<td></td>
</tr>
<tr>
<td>View the following website for plagiarism and APA guidelines: <a href="http://fpdc.kent.edu/PlagiarismAPA.htm">http://fpdc.kent.edu/PlagiarismAPA.htm</a></td>
<td></td>
</tr>
<tr>
<td>Write first draft of paper, making sure to follow APA format</td>
<td></td>
</tr>
<tr>
<td>Submit first draft of paper by <strong>March 2, 11:59 pm</strong>.</td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>March 10</strong></td>
<td></td>
</tr>
<tr>
<td>Prepare for reviewing by viewing the following website for examples of reviews: <a href="http://www.peerfeedback.net">www.peerfeedback.net</a>. Complete accompanying question sheet labeled “Peerfeedback.net” and submit to instructor.</td>
<td></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>March 9, 11:59 pm</strong>.</td>
<td></td>
</tr>
<tr>
<td>Submit Second Draft by <strong>March 16, 11:59 pm</strong>.</td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft by <strong>March 23, 11:59 pm</strong>. <strong>Do not back review anyone with blank comment fields!</strong></td>
<td></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>April 6, 11:59 pm</strong>.</td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>April 13</strong>.</td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft by <strong>April 13, 11:59 pm.</strong> <strong>Do not back review anyone with blank comment fields!</strong></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX I

Group 1 Instructions
Instructions for Educational Psychology Theorist Research Paper

Overview

For this assignment, you are asked to write two drafts of a 5-6 page research paper on an educational psychology learning theory or approach to teaching and learning. In addition to producing 2 drafts of your paper, you are asked to review 5-6 peer papers for each draft. You will be evaluated on both your writing and your reviewing as described below in the evaluation section.

Attached to this assignment sheet is a list of topics that should guide your selection. You should collect relevant information for your topic that is most valuable to teachers in training. You should select an approach to teaching and learning, collect scholarly sources that best address your topic (at least four) and write a detailed explanation of the particular approach to education, tying theory to current classroom practice. In order to avoid plagiarism, you must cite (document) the sources of your work in the text of your paper and list all of your sources in an APA reference list at the end of your paper. Online reference guides for APA and plagiarism are available at the following web address: http://fpdc.kent.edu/PlagiarismAPA.htm

Due Dates:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submit First Draft</td>
<td>2/9/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>3/9/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Second Draft</td>
<td>3/16/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft</td>
<td>3/23/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Review Assigned Peer Papers</td>
<td>4/6/06</td>
<td>Midnight</td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft</td>
<td>4/13/06</td>
<td>Midnight</td>
</tr>
</tbody>
</table>

Instructions:

After selecting a topic and clearing it with your instructor, the first thing to do is to register on the SWoRD system. For this, you will need a computer with an internet connection as well as the SWoRD Student Manual. It is important to do this at least a week prior to the first deadline in order to deal with problems that may arise. To do this, have open your SWoRD Student Manual and using your web browser, navigate to http://ladybug.lrdc.pitt.edu/sword3. From here, follow your student manual page by page. When asked to select a course, select “2006 Spring (RWooley)Educational Psychology Kent State University”. When asked to select a topic, choose Topic One – Red for writing and Topic 2 – Blue for Reviewing. Be careful to select the right topics! Notice that you are selecting a different topic for writing than for reviewing.

After registering on SWoRD, read the rest of the manual in order to understand the entire process. You will need to come back to this section of the manual when you are ready to begin submitting assignments. Do not wait until the last minute to submit assignments! Read section 5, “SWoRD Policy” on page 20.
Use the checklist on the following page to help guide your progress on the assignment. You are required to complete every step on the checklist, including looking at example files as assigned. The checklist is very, very important.
<table>
<thead>
<tr>
<th>Task</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read over assignment sheets and instructions</td>
<td></td>
</tr>
<tr>
<td>Choose a topic from list and clear it with instructor</td>
<td></td>
</tr>
<tr>
<td>Use SWoRD Student Manual and <a href="http://ladybug.lrdc.pitt.edu/sword3">link</a> to register for the SWoRD system by <strong>January 27</strong>.</td>
<td></td>
</tr>
<tr>
<td>View the following website for plagiarism and APA guidelines: <a href="http://fpdc.kent.edu/PlagiarismAPA.htm">link</a></td>
<td></td>
</tr>
<tr>
<td>Write first draft of paper, making sure to follow APA format</td>
<td></td>
</tr>
<tr>
<td>Submit first draft of paper by <strong>February 9, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>February 17</strong></td>
<td></td>
</tr>
<tr>
<td>Prepare for reviewing by viewing the following website for examples of reviews: <a href="www.peerfeedback.net">link</a>. Complete accompanying question sheet labeled “Peerfeedback.net” and submit to instructor.</td>
<td></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>March 9, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Second Draft by <strong>March 16, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for First Draft by <strong>March 23, 11:59 pm.</strong></td>
<td><strong>Do not back review anyone with blank comment fields!</strong></td>
</tr>
<tr>
<td>Review Assigned Peer Papers by <strong>April 6, 11:59 pm.</strong></td>
<td></td>
</tr>
<tr>
<td>Check scores in SWoRD to gauge progress in class <strong>April 13.</strong></td>
<td></td>
</tr>
<tr>
<td>Submit Back Reviews for Second Draft by <strong>April 13, 11:59 pm.</strong></td>
<td><strong>Do not back review anyone with blank comment fields!</strong></td>
</tr>
</tbody>
</table>
APPENDIX J

Consent Form
January 22, 2006

Prospective Participant:

I would like to do research on the effects of reviewing on student writing using an online peer review system. I would like to do this because it will contribute knowledge to the field of educational psychology by shedding light on writing and learning processes. Because this study makes use of an existing course assignment, you are required to complete the assignment by your instructors. If you agree to participate in this study, you are simply giving me permission to use your scores anonymously in the compilation of results. There is no advantage or disadvantage to participating in the study. Your identity during and after the study will be kept confidential.

Your assignment will be submitted to an electronic peer review system as described in your packet of instructions for the assignment. The system requires you to create a pseudonym (nickname) which becomes the only way your work is identified to other students and to the researcher.

If you agree to participate in this study, you will be contributing to the advancement of knowledge about teaching and learning. Allowing your scores to be used in this research project is entirely up to you, and no one will hold it against you if you decide not to do it. If you do take part, you may change your mind at any time.

If you want to know more about this project, please call me at 330-672-7840 or call my advisor, Dr. Dalton, at 330-672-9079. The project has been approved by Kent State University. If you have questions about Kent State University’s rules for research, please call Dr. John L. West, Vice President and Dean, Division of Research and Graduate Studies (330-672-2704).

You will get a copy of this consent form.

Sincerely,

Ryan S. Wooley
Assistant Director,
Faculty Professional Development Center

CONSENT STATEMENT

I agree to take part in this project. I understand what I have to do and that I can change my mind at any time.

Signature

Date
APPENDIX K

Schedule Overview
### Timeline for SWoRD Assignment

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wooley Explains assignment to classes</td>
<td>Submit Draft 1</td>
<td>Review Draft 1 for groups 2-6</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wooley Explains assignment to classes</td>
<td>Review Draft 1 for groups 2-6</td>
<td>Submit Draft 1</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Wooley Explains assignment to classes</td>
<td>Submit Draft 1</td>
<td>Submit Draft 1</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Wooley Explains assignment to classes</td>
<td>Submit Draft 1</td>
<td>Submit Draft 1</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Wooley Explains assignment to classes</td>
<td>Submit Draft 1</td>
<td>Submit Draft 1</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Wooley Explains assignment to classes</td>
<td>Submit Draft 1</td>
<td>Review Draft 1 for groups 2-6</td>
<td>Submit Draft 2</td>
<td>Submit Back Reviews for Draft 1</td>
<td>Spring Break</td>
<td>Review Assigned Papers</td>
<td>Submit Back Reviews for Draft 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Group Size (N) Treatment

<table>
<thead>
<tr>
<th>Group</th>
<th>Size (N)</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>Write First Group, Not included in study. Provides review material for groups 2-5</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>Review First, Rating Only, Without Examples</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>Review First, Rating Only, With Examples</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>Review First, Rate and Comment, Without Examples</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>Review First, Rate and Comment, With Examples</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>Write First, appended control group</td>
</tr>
</tbody>
</table>
APPENDIX L

Logistical Design
N=65  \textit{WE/RL}  \\
N=70  \textit{RE/WL}  \\
N=70  \textit{WL/RL}  \\

* Nested Factorial, with N=17 for each group. Factors S/NS, and RO, RC defined below.

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>S = With samples of helpful and unhelpful reviews</td>
<td>Students are given written instructions and examples for feedback</td>
</tr>
<tr>
<td>NS = No samples of helpful and unhelpful reviews</td>
<td>Sans above</td>
</tr>
<tr>
<td>RO = Rating only (no comments)</td>
<td>Reviewing and Commenting on papers</td>
</tr>
<tr>
<td>RC = Rating + comments</td>
<td>Students see other papers but do not comment on them.</td>
</tr>
<tr>
<td>WE/RL = Write Early, Review Late</td>
<td>Students in this group will write their first draft early, and prior to reviewing.</td>
</tr>
<tr>
<td>RE/WL = Review Early, Write Late</td>
<td>Students in this group will review the WE/RL group drafts prior to writing.</td>
</tr>
<tr>
<td>WL/RL = Write Late, Review Late</td>
<td>Students in this group will write at the same time as RE/RL, but without reviewing first</td>
</tr>
<tr>
<td>lag 1 and lag 3</td>
<td>One week</td>
</tr>
<tr>
<td>lag 2</td>
<td>Two weeks</td>
</tr>
<tr>
<td>W1, W2, R1, R2</td>
<td>Respectively—Draft 1, Draft 2, Review 1, Review 2</td>
</tr>
<tr>
<td>Arrows</td>
<td>Designate intergroup assignment/distribution of drafts for reviewing</td>
</tr>
</tbody>
</table>
APPENDIX M

Assignment Sheet - Examples Treatment
1. Find a computer with Internet access and an ability to play a video with sound that you can hear.
2. Go to http://www.peerfeedback.net/
3. Run through the presentation (click on Launch), keeping the following questions in mind.
4. Answer the following questions and hand in the answers prior to your first review date.
   a. What are main points made in the video? ____________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
   b. Which points did you find obvious (if any)? ________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
   c. Which points did you find surprising (if any)? ______________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
   d. What other questions do you have about giving feedback to peers? _______
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
      ________________________________________________________________
APPENDIX N

Adjusted Group Results
Adjusted Group Results

A one-sample Komogorov-Smirnov test revealed that data had adequate normality, $Z=1.13$, $p=.15$. Levene’s test of equality of error variances reveal that homogeneity of variance was not tenable, $F(3,51) = 5.10$, $p = .00$.

Hypothesis 1

Hypothesis 1 predicted that mean writing scores in the with-examples treatment group would be higher than those in the without-examples group. Because this treatment did not involve competing sets of instructions, such as print instructions directing students to ignore online instructions to make elaborate comments, it was not as likely for treatments to be mixed up. Further, the extent to which this treatment could be tracked was much less precise than for the elaboration treatments, where the SWoRD database served as a record of student comments. Therefore, the examples group could not be adjusted in any way, and results are the same as shown in Chapter IV. These results are repeated here: As shown in Table N1, mean differences were in the opposite direction from those predicted (with-examples, $M = 43.71$, $SD = 17.72$; without-examples, $M = 50.50$, $SD = 20.61$). As shown in Table N2, these differences were not statistically significant: $F(1,55) = 2.18$, $p = .15$. These results do not support the hypothesis that providing examples of prototypical helpful and unhelpful reviews would lead to higher quality of reviewers’ subsequent writing.
**Hypothesis 2**

Hypothesis 2 predicted that mean writing scores in the with-elaboration treatment group would be higher than those in the without-elaboration group. As shown in Table N1, mean differences were consistent with the predicted direction (with-elaboration, $M = 52.30, SD = 20.60$; without-elaboration, $M = 36.24, SD = 10.17$). As shown in Table N2, these differences were statistically significant: $F(1,55) = 10.69, p = .00$. Partial Eta$^2$ was .17. These results support the hypothesis that providing elaborate comments in addition to simple numeric ratings would lead to higher quality of reviewers’ subsequent writing.

**Hypothesis 3**

Hypothesis 3 predicted an interaction between examples and elaboration treatments. As shown in Table N1, means are ordered consistently across treatment groups. Each without-examples group mean is higher than its respective with-examples mean, and each with-elaboration group mean is higher than its respective without-elaboration group mean (without-examples, with-elaboration $M = 57.23, SD = 22.81$; without-examples, without-elaboration, $M = 39.04, SD = 8.39$; with-examples, with-elaboration $M = 48.10, SD = 18.03$; with-examples, without-elaboration, $M = 32.74, SD = 11.64$). As shown in Table N2, these differences were not statistically significant: $F(1,55) = .08, p = .78$. These results do not support the hypothesis that there would be an interaction between examples and elaboration treatments.
### Table N1
*Adjusted Score Means and Standard Deviations by Factorial Cell*

<table>
<thead>
<tr>
<th></th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Only</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M = 32.74$</td>
<td>$M = 39.04$</td>
<td>$M = 36.24$</td>
<td></td>
</tr>
<tr>
<td>$SD = 11.64$</td>
<td>$SD = 8.39$</td>
<td>$SD = 10.17$</td>
<td></td>
</tr>
<tr>
<td>$n = 8$</td>
<td>$n = 10$</td>
<td>$n = 18$</td>
<td></td>
</tr>
<tr>
<td><strong>Rating and</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Commenting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M = 48.10$</td>
<td>$M = 57.23$</td>
<td>$M = 52.30$</td>
<td></td>
</tr>
<tr>
<td>$SD = 18.03$</td>
<td>$SD = 22.81$</td>
<td>$SD = 20.60$</td>
<td></td>
</tr>
<tr>
<td>$n = 20$</td>
<td>$n = 17$</td>
<td>$n = 37$</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M = 43.71$</td>
<td>$M = 50.50$</td>
<td>$M = 47.04$</td>
<td></td>
</tr>
<tr>
<td>$SD = 17.72$</td>
<td>$SD = 20.61$</td>
<td>$SD = 19.32$</td>
<td></td>
</tr>
<tr>
<td>$n = 28$</td>
<td>$n = 27$</td>
<td>$n = 55$</td>
<td></td>
</tr>
</tbody>
</table>
### Table N2

**Adjusted Score Factorial ANOVA - Main Effects and Interactions**

<table>
<thead>
<tr>
<th></th>
<th>MS</th>
<th>df</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples</td>
<td>714.09</td>
<td>1</td>
<td>2.26</td>
<td>.04</td>
<td>.14</td>
</tr>
<tr>
<td>Elaboration</td>
<td>3372.15</td>
<td>1</td>
<td>10.69</td>
<td>.17</td>
<td>.00</td>
</tr>
<tr>
<td>Interaction</td>
<td>23.89</td>
<td>1</td>
<td>.07</td>
<td>.00</td>
<td>.78</td>
</tr>
<tr>
<td>Error</td>
<td>315.37</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary**

These results are consistent with the results of original group analyses as reported in Chapter IV, without group adjustments. The results are not supportive of the hypothesis of Phase I: that articulation of feedback by reviewers using a web-based peer review system for writing would result in higher quality of reviewers’ subsequent writing. For Phase II, results were not supportive of hypothesis 1—that exposure to prototypical examples of feedback during reviewing would result in better subsequent writing. Hypothesis 2—that students who provided more elaborate feedback (ratings and comments) would perform better on the writing assignment than students who provided ratings alone—was supported. Finally, hypothesis 3, that there would be an interaction between examples and elaboration treatments, was unsupported.
APPENDIX O

SWoRD Score Results
SWoRD Score Results

A one-sample Komogorov-Smirnov test revealed that data had adequate normality, $Z = .75, p = .62$. Levene’s test of equality of error variances revealed that homogeneity of variance was tenable, $F(3,51) = .34, p = .80$.

**Hypothesis 1**

Hypothesis 1 predicted that mean writing scores in the with-examples treatment group would be higher than those in the without-examples group. As shown in Table O1, mean differences were in the order predicted (with-examples, $M = 80.52, SD = 7.41$; without-examples, $M = 78.86, SD = 6.17$). As shown in Table O2, these differences were not statistically significant: $F(1,55) = .79, p = .38$. These results do not support the hypothesis that providing examples of prototypical helpful and unhelpful reviews would lead to higher quality of reviewers’ subsequent writing.

**Hypothesis 2**

Hypothesis 2 predicted that mean writing scores in the with-elaboration treatment group would be higher than those in the without-elaboration group. As shown in Table O1, mean differences were consistent with the predicted direction (with-elaboration, $M = 80.15, SD = 6.98$; without-elaboration, $M = 79.28, SD = 6.77$). As shown in Table O2, these differences were not statistically significant: $F(1,55) = .20, p = .66$. These results do not support the hypothesis that providing elaborate comments in addition to simple numeric ratings would lead to higher quality of reviewers’ subsequent writing.
Hypothesis 3

Hypothesis 3 predicted an interaction between examples and elaboration treatments. As shown in Table O1, means are ordered consistently across treatment groups. Each with-examples group mean is higher than its respective without-examples mean, and each with-elaboration group mean is higher than its respective without-elaboration group mean (without-examples, with-elaboration $M = 78.92$, $SD = 5.84$; without-examples, without-elaboration, $M = 78.80$, $SD = 6.69$; with-examples, without-elaboration, $M = 81.29$, $SD = 7.94$; with-examples, without-elaboration, $M = 79.75$, $SD = 7.06$). As shown in Table O2, these differences were not statistically significant: $F(1,55) = .14$, $p = .71$. These results do not support the hypothesis that there would be an interaction between examples and elaboration treatments.
Table O1
SWoRD Score Means and Standard Deviations by Factorial Cell

<table>
<thead>
<tr>
<th></th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating Only</td>
<td>$M = 79.75$</td>
<td>$M = 78.80$</td>
<td>$M = 79.28$</td>
</tr>
<tr>
<td></td>
<td>$SD = 7.06$</td>
<td>$SD = 6.69$</td>
<td>$SD = 6.77$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 14$</td>
<td>$n = 28$</td>
</tr>
<tr>
<td>Rating and Commenting</td>
<td>$M = 81.29$</td>
<td>$M = 78.92$</td>
<td>$M = 80.15$</td>
</tr>
<tr>
<td></td>
<td>$SD = 7.94$</td>
<td>$SD = 5.84$</td>
<td>$SD = 6.98$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 13$</td>
<td>$n = 27$</td>
</tr>
<tr>
<td>Total</td>
<td>$M = 80.52$</td>
<td>$M = 78.86$</td>
<td>$M = 79.71$</td>
</tr>
<tr>
<td></td>
<td>$SD = 7.41$</td>
<td>$SD = 6.17$</td>
<td>$SD = 6.82$</td>
</tr>
<tr>
<td></td>
<td>$n = 28$</td>
<td>$n = 27$</td>
<td>$n = 55$</td>
</tr>
</tbody>
</table>
Summary

These results do not support any of the hypotheses of Phase II, however, as described in Chapter V, SWoRD scores in this context are suspect and were discarded in favor of expert scores. Students, concerned about their review grades, tended to score toward the center of the Likert scale, making few distinctions between papers. This tendency was exacerbated by the context of the investigation. The high kurtosis among SWoRD scores shown in Chapter V, Figure 1 supports these observations.
APPENDIX P

MANOVA Results
MANOVA Results

Phase I

The research hypothesis for Phase I was that articulation of feedback by reviewers using a web-based peer review system for writing will affect the quality of reviewers’ subsequent writing. The hypothesis was analyzed through a single factor MANOVA with three dependent variables: (a) flow, (b) logic and (c) insight. Results indicate no significant differences for any of the dependent variables based on this treatment. MANOVA assumptions were tenable, except that Box’s Test of Equality of Covariance Matrices revealed that covariance matrices of the dependent variables were not equal across groups ($p=.04$). However, violations of this assumption are thought to have minimal impact when group sizes are approximately equal (Hair, Anderson, & Black, 1998), as they are in the current study. Table P1 presents the means and standard deviations for writing scores in each dependent variable by treatment group. Mean differences are quite small, varying at most .17 from their respective grand means. It is worth noting that members of the control group outscored members of the treatment group in all three dimensions, though not at a statistically significant level.
The one-factor MANOVA examining effects of reviewing indicates no significant differences between treatment and control groups, Wilk’s $\lambda = .97$, $F (1,115) = .36$, $p = .20$.

**Phase II**

*Preliminary analyses.*

Preliminary analyses revealed that a few students had markedly low scores on the assignment. Upon examination of student artifacts for these students, it was discovered that three students had submitted assignments too incomplete to be evaluated by the dimensional criteria. For example, one student submitted a single-page bulleted list, when the assignment was to write a 5-6 page paper. Based on this discovery, three students were dropped from the analysis.

While MANOVA assumptions of equality of covariance matrices and equality of error variance across groups were not tenable, the F-statistic is thought to be robust in the
face of violations of these assumptions, especially when group sizes are approximately equal (Glass, Peckham, & Sanders, 1972; Hair, Anderson, & Black, 1998). According to Stevens (1996), the assumption of equal covariance matrices is restrictive and rarely satisfied in practice. He also notes that the Box Test (1949), the most common means of testing homogeneity of covariance matrices, is very sensitive to non-normality. Stevens suggests transforming data for even marginal departures from normality prior to employing the Box Test. In the current study, when original scores are transformed logarithmically, nearly all assumptions, including homogeneity of covariance matrices are satisfied. Since MANOVA results on these transformed scores are nearly identical to results obtained from original scores, original scores were retained in subsequent analyses.

Because the rationale for MANOVA is largely based on correlations between dependent variables, Pearson product-moment correlations were calculated. Table P2 presents the correlations between the dependent variables, and shows that dependent variables are highly correlated.

<table>
<thead>
<tr>
<th>Table P2</th>
<th>Pearson Correlation Coefficients for Dependent Variables – Phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension/Dimension</strong></td>
<td><strong>Flow</strong></td>
</tr>
<tr>
<td>Flow</td>
<td>---</td>
</tr>
<tr>
<td>Logic</td>
<td><em>r</em> = .820</td>
</tr>
<tr>
<td>Insight</td>
<td><em>r</em> = .911</td>
</tr>
</tbody>
</table>
After preliminary analyses, the following research hypotheses were tested in Phase II:

1. The provision of examples of helpful and unhelpful reviews to reviewers using a web-based peer review system for writing will affect the quality of reviewers’ subsequent writing.

2. Providing elaborate comments in addition to reviewer ratings will affect the quality of reviewers’ subsequent writing.

3. There will be an interaction between example exposure type and feedback elaboration type.

These hypotheses were analyzed through a 2 x 2 factorial MANOVA, where the independent variables were: (a) use of examples to guide student review (included versus excluded) and (b) level of feedback elaboration required by reviewers (elaborate versus simple), and the dependent variables are (a) flow, (b) logic and (c) insight.

Tables P3, P4, and P5 present the means and standard deviations for writing scores in each dependent variable by factorial cell.

*Hypothesis 1.*

Hypothesis 1 predicted that mean writing scores in the with-examples treatment group would be higher than those in the without-examples group. As shown in Table P3, mean differences on the flow dimension were in the opposite direction from those predicted (with-examples, $M = 3.46$, $SD = 1.19$; without-examples, $M = 3.73$, $SD = 1.43$). Shown in Table P4, mean differences on the logic dimension were in the opposite direction from those predicted (with-examples, $M = 2.89$, $SD = 1.27$; without-examples, $M = 3.28$, $SD = 1.52$). Shown in Table P5, mean differences on the insight dimension
were in the opposite direction from those predicted (with-examples, $M = 2.82$, $SD = 1.34$; without-examples, $M = 3.59$, $SD = 1.66$). As shown in Table P6, these differences were statistically significant: $F(3,49) = 3.58$, $p = .02$. These results do not support the hypothesis that providing examples of prototypical helpful and unhelpful reviews would lead to higher quality of reviewers’ subsequent writing on any of the three writing dimensions.

**Hypothesis 2.**

Hypothesis 2 predicted that mean writing scores in the with-elaboration treatment group would be higher than those in the without-elaboration group. As shown in Table P3, mean differences on the flow dimension were consistent with the prediction (with-elaboration, $M = 4.14$, $SD = 1.48$; without-elaboration, $M = 3.07$, $SD = .085$). Shown in Table P4, mean differences on the logic dimension were also consistent with predictions (with-elaboration, $M = 3.58$, $SD = 1.60$; without-elaboration, $M = 2.67$, $SD = 0.99$). Shown in Table P5, mean differences on the insight dimension were also consistent with predictions (with-elaboration, $M = 3.79$, $SD = 1.73$; without-elaboration, $M = 2.63$, $SD = 1.10$). As shown in Table P6, these differences were also statistically significant: $F(3,49) = 3.50$, $p\eta^2 = .18$, $p = .02$. These results support the hypothesis that providing elaborate comments in addition to simple numeric ratings would lead to higher quality of reviewers’ subsequent writing.

**Hypothesis 3.**

Hypothesis 3 predicted an interaction between examples and elaboration treatments. As shown in Tables P3, P4, and P5, means are ordered consistently across
treatment groups. Each without-examples group mean is higher than its respective with-examples mean, and each with-elaboration group mean is higher than its respective without-elaboration group mean. As shown in Table P6, these differences were not statistically significant: \(F(3,49) = .54, p = .66\). These results do not support the hypothesis that there would be an interaction between examples and elaboration treatments.

**Table P3**  
*Means and Standard Deviations for Flow Dimension*

<table>
<thead>
<tr>
<th></th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Only</strong></td>
<td>(M = 3.05)</td>
<td>(M = 3.10)</td>
<td>(M = 3.07)</td>
</tr>
<tr>
<td></td>
<td>(SD = 0.95)</td>
<td>(SD = 0.78)</td>
<td>(SD = 0.85)</td>
</tr>
<tr>
<td></td>
<td>(n = 14)</td>
<td>(n = 14)</td>
<td>(n = 28)</td>
</tr>
<tr>
<td><strong>Rating and Commenting</strong></td>
<td>(M = 3.88)</td>
<td>(M = 4.41)</td>
<td>(M = 4.14)</td>
</tr>
<tr>
<td></td>
<td>(SD = 1.29)</td>
<td>(SD = 1.67)</td>
<td>(SD = 1.48)</td>
</tr>
<tr>
<td></td>
<td>(n = 14)</td>
<td>(n = 13)</td>
<td>(n = 27)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>(M = 3.46)</td>
<td>(M = 3.73)</td>
<td>(M = 3.59)</td>
</tr>
<tr>
<td></td>
<td>(SD = 1.19)</td>
<td>(SD = 1.43)</td>
<td>(SD = 1.30)</td>
</tr>
<tr>
<td></td>
<td>(n = 28)</td>
<td>(n = 27)</td>
<td>(n = 55)</td>
</tr>
</tbody>
</table>
Table P4  
*Means and Standard Deviations for Logic Dimension*

<table>
<thead>
<tr>
<th></th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Only</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>2.41</td>
<td>2.81</td>
<td>2.67</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.92</td>
<td>1.04</td>
<td>0.99</td>
</tr>
<tr>
<td>$n$</td>
<td>14</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td><strong>Rating and</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Commenting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>3.88</td>
<td>3.80</td>
<td>3.58</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.29</td>
<td>1.81</td>
<td>1.60</td>
</tr>
<tr>
<td>$n$</td>
<td>14</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>2.89</td>
<td>3.28</td>
<td>3.08</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.27</td>
<td>1.52</td>
<td>1.40</td>
</tr>
<tr>
<td>$n$</td>
<td>28</td>
<td>27</td>
<td>55</td>
</tr>
</tbody>
</table>
### Table P5
**Means and Standard Deviations for Insight Dimension**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>With Examples</th>
<th>Without Examples</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating Only</strong></td>
<td>$M = 2.29$</td>
<td>$M = 2.98$</td>
<td>$M = 2.63$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.08$</td>
<td>$SD = 1.04$</td>
<td>$SD = 1.10$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 14$</td>
<td>$n = 28$</td>
</tr>
<tr>
<td><strong>Rating and</strong></td>
<td>$M = 3.36$</td>
<td>$M = 4.26$</td>
<td>$M = 3.79$</td>
</tr>
<tr>
<td><strong>Commenting</strong></td>
<td>$SD = 1.40$</td>
<td>$SD = 1.97$</td>
<td>$SD = 1.73$</td>
</tr>
<tr>
<td></td>
<td>$n = 14$</td>
<td>$n = 13$</td>
<td>$n = 27$</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$M = 2.82$</td>
<td>$M = 3.59$</td>
<td>$M = 3.20$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.34$</td>
<td>$SD = 1.66$</td>
<td>$SD = 1.54$</td>
</tr>
<tr>
<td></td>
<td>$n = 28$</td>
<td>$n = 27$</td>
<td>$n = 55$</td>
</tr>
</tbody>
</table>

The descriptive statistics in Tables P3, P4, and P5 show a consistent order of means across each of the dependent variables for elaboration treatments, with with-elaboration group means higher than the without-elaboration group means in flow, logic and insight. These results are consistent with the hypothesis that providing elaborate comments in addition to simple numerical ratings would result in higher quality of reviewers’ subsequent writing. Tables P3, P4, and P5 also show a consistent order of means across the dimensions for examples treatments, with the with-examples group means lower than
the without-examples group means in flow, logic and insight. These results are not consistent with the hypothesis that providing examples of prototypical feedback would result in higher quality of reviewers’ subsequent writing. While this finding is inconsistent with predictions that providing examples of prototypical reviews would result in subsequent higher quality writing by reviewer, effects were significant in the opposite direction from predictions.

As shown in Table P6, the MANOVA summary reveals effects for both main independent variables, but not for the interaction. The effect size was the same across the examples ($\eta^2 = .18$) and elaboration ($\eta^2 = .18$) treatments.

<table>
<thead>
<tr>
<th>Table P6</th>
<th>Factorial MANOVA - Main Effects and Interactions for Examples and Elaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$</td>
</tr>
<tr>
<td>Examples</td>
<td>.82</td>
</tr>
<tr>
<td>Elaboration</td>
<td>.82</td>
</tr>
<tr>
<td>Interaction</td>
<td>.97</td>
</tr>
</tbody>
</table>

These results are not supportive of the first hypothesis of Phase II: that exposure to prototypical examples of feedback during reviewing would result in better subsequent writing. Exposure to examples did have a significant effect on writing performance, but in exactly the opposite direction from what was predicted; exposure to examples resulted in poorer quality writing. However, the second hypothesis is supported by the MANOVA
results: Students who provided more elaborate feedback (ratings and comments)
performed better than students who provided ratings alone.

After initial MANOVA showed significant effects for both main treatment
variables, successive MANOVAs were performed to determine dependent variable
contribution (Wilkinson, 1975). As discussed in detail in Chapter III, this type of post
hoc analysis solves many of the problems associated with the more popular method of
running multiple univariate ANOVAs – most importantly that separate ANOVAs ignore
correlations between the dependent variables (Bray & Maxwell, 1982). Dependent
variable contribution is determined by running additional successive MANOVAs, each
time dropping a specific dependent variable from analysis. Differences in effect between
one of these post hoc MANOVAs and the original all-inclusive MANOVA reveal the
contribution of the dropped dependent variable.

As shown in Table P7, when the influence of flow is removed from the original
MANOVA, under the examples factor, $F(3,49) = 3.58$ changes to $F(2,50) = 2.93$, and $p\eta^2$
= .18 changes to $p\eta^2 = .11$. Thus, flow makes a .07 contribution to the effect of the
examples factor, and when it is removed, significance drops below an acceptable value.
Also shown in Table P7, when flow is removed from the MANOVA, under the
elaboration factor, $F(3,49) = 3.50$ changes to $F(2,50) = 4.71$, and $p\eta^2 = .18$ changes to $p\eta^2$
= .16. Thus, flow makes a .02 contribution to the effect of the examples factor.
As shown in Table P8, when the influence of logic is removed from the original MANOVA, under the examples factor, $F(3, 49) = 3.58$ changes to $F(2, 50) = 4.90$, and $\eta^2 = .18$ changes to $\eta^2 = .16$. Thus, logic makes a .02 contribution to the effect of the examples factor. Also shown in Table P8, when logic is removed from the MANOVA, under the elaboration factor, $F(3, 49) = 3.50$ changes to $F(2, 50) = 5.36$, and $\eta^2 = .18$ remains the same. Thus, logic makes no contribution to the effect of the examples factor.

### Table P7

*Factorial MANOVA - Main Effects for Examples and Elaboration without Flow*

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples Original</td>
<td>.82</td>
<td>3,49</td>
<td>3.58</td>
<td>.02</td>
<td>.18</td>
</tr>
<tr>
<td>Examples without Flow</td>
<td>.90</td>
<td>2,50</td>
<td>2.93</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td>Elaboration Original</td>
<td>.82</td>
<td>3,49</td>
<td>3.50</td>
<td>.02</td>
<td>.18</td>
</tr>
<tr>
<td>Elaboration without Flow</td>
<td>.84</td>
<td>2,50</td>
<td>4.71</td>
<td>.01</td>
<td>.16</td>
</tr>
</tbody>
</table>

### Table P8

*Factorial MANOVA - Main Effects for Examples and Elaboration without Logic*

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples Original</td>
<td>.82</td>
<td>3,49</td>
<td>3.58</td>
<td>.02</td>
<td>.18</td>
</tr>
<tr>
<td>Examples without Logic</td>
<td>.84</td>
<td>2,50</td>
<td>4.90</td>
<td>.01</td>
<td>.16</td>
</tr>
<tr>
<td>Elaboration Original</td>
<td>.82</td>
<td>3,49</td>
<td>3.50</td>
<td>.02</td>
<td>.18</td>
</tr>
<tr>
<td>Elaboration without Logic</td>
<td>.82</td>
<td>2,50</td>
<td>5.36</td>
<td>.01</td>
<td>.18</td>
</tr>
</tbody>
</table>
As shown in Table P9, when the influence of insight is removed from the original MANOVA, under the examples factor, $F(3,49) = 3.58$ changes to $F(2,50) = .64$, and $\eta^2 = .18$ changes to $\eta^2 = .03$. Thus, insight makes a .15 contribution to the effect of the examples factor. It might be noted as well that, when the influence of the logic factor is removed, significance changes markedly as well—going from $p = .02$ to $p = .53$. Also shown in Table P9, when insight is removed from the MANOVA, under the elaboration factor, $F(3,49) = 3.50$ changes to $F(2,50) = 5.34$, and $\eta^2 = .18$ changes to $\eta^2 = .18$. Thus, insight makes no contribution to the effect of the examples factor.

<p>| Table P9 |</p>
<table>
<thead>
<tr>
<th>Factorial MANOVA - Main Effects for Examples and Elaboration without Insight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
</tr>
<tr>
<td>Examples Original</td>
</tr>
<tr>
<td>Examples without Insight</td>
</tr>
<tr>
<td>Elaboration Original</td>
</tr>
<tr>
<td>Elaboration without Insight</td>
</tr>
</tbody>
</table>

These results show that insight made the strongest, and only significant contribution to the overall effect of examples treatments. In other words, the insight dimension in particular saw significant negative effects from examples treatments. These results also indicate that dependent variables made fairly equal contributions toward the effects of elaboration treatments.
REFERENCES
REFERENCES


Cennamo, K. S. (1989, February). Factors influencing mental effort: A theoretical overview and review of literature. Presented at the annual meeting of the Association for Educational Communication and Technology, Dallas, TX.


Intersegmental Committee of the Academic Senates of the California Community Colleges, the California State University, and the University of California. (2002). *Academic literacy: A statement of competencies expected of students entering California’s public colleges and universities*. Sacramento: ICAS.


explanations in computer-based learning environments. *Interactive Learning
Environments, 10*(2), 105-119.

presented at annual Conference on College Composition and Communication,
Atlanta.


Salomon, G. (1983). The differential investment of mental effort in learning from


reciprocal evaluation*. PowerPoint presentation delivered at the October Campus
Conversations event Kent State University, Kent, OH.

individual knowledge in argumentative activity. *The Journal of the Learning
Sciences, 12*(2), 219-256.


Shaw, V. N. (2002). Peer review as a motivating device in the training of writing skills for college students. *Journal of College Reading and Learning, 33*(1), 68-76.


Simmons, J. (2003). Responders are taught, not born. *Journal of Adolescent and Adult Literacy, 46*(8), 684-693.


