PEER REVIEW OF WRITING: LEARNING FROM REVISION USING PEER FEEDBACK AND REVIEWING PEERS' TEXTS

by

Melissa M. Patchan

B.A., University of South Florida, 2004

M.S., University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the College of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Melissa M. Patchan

It was defended on

January 12, 2011

and approved by

Kevin D. Ashley, Professor of Law

Christine M. Neuwirth, Professor of English

Timothy J. Nokes, Assistant Professor of Psychology

Dissertation Advisor: Christian D. Schunn, Associate Professor of Psychology

PEER REVIEW OF WRITING:

LEARNING FROM REVISION USING PEER FEEDBACK

AND REVIEWING PEERS' TEXTS

Melissa M. Patchan, PhD

University of Pittsburgh, 2011

To better understand the ways in which peer feedback can vary across contexts to shape learning outcomes, the current dissertation examined how student ability levels shapes peer review. Peer review involves opportunities to learn from receiving feedback and providing feedback. Both elements were considered within one larger experiment. A sample of 189 students enrolled in an Introduction to Psychological Science course participated in this experiment. As part of the requirements for class, students wrote a first draft of their paper, reviewed four peers' papers, revised their draft based on peer feedback, and wrote a new paper on a different topic. In this experiment, I manipulated to whom the students provided feedback and from whom they received feedback in order to observe the relationships between writer ability, quality of text reviewed, and ability of the feedback source. In the first study, I determined which types of comments had an impact on draft quality and whether there were differences in the amounts received and implemented depending on the writer ability and feedback source. In the second study, I examined the extent to which students could learn from reviewing peers' texts and how the ability of the writer and the quality of the text being reviewed influenced the amount of improvement. Based on the findings from the two studies, theoretical and pragmatic implications were offered.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 GENERAL INTRODUCTION

Many instructors are hesitant to require peer review of writing—they are most concerned whether low ability students are capable of helping other students and whether high-ability students could learn anything from reviewing low-ability students' papers. To address these concerns, the current dissertation examined how students of varying abilities learn from the peer review process. Specifically, two questions were considered: 1) how does the student's writing ability and the writing ability of the feedback source affect the revision process—the types of peer feedback received, the implementability of the peer feedback, and the quality of the revisions made based on the peer feedback, and 2) how does the student's writing ability and the quality of the text to be reviewed affect how much students improve their writing ability from the various activities involved in reviewing peers' texts—reading the text, evaluating the quality of the paper, and detecting and diagnosing problems. From a better understanding of how students revise their own papers using peer feedback and improve their writing ability from reviewing peers' texts, instructors could develop peer review assignments that maximize learning benefits for students overall and for particular subgroups of students within their classes.

Many researchers have examined the effects of peer review on student writing. However, these studies were limited in a number of ways: researchers only focused on the effects of receiving peer feedback or providing feedback to peers, classroom assignments needed to be modified in a way that was only necessary for the experiment, the quality of revisions were not measured quantitatively, and researchers did not observe the effects on future writing. In the current dissertation, new methods were used to address these limitations.

First, several researchers who conducted more rigorous experiments to further explain the observed effects used experimental designs that disrupted the natural flow of the course. For example, Wooley, Was, Schunn & Dalton (2008) examined whether students wrote higher quality papers after reviewing peers' texts (i.e., only rating the quality of the paper vs. also providing comments). To compare students who reviewed peers' texts to those who had not reviewed peers' texts, some students needed to complete their draft earlier than the remaining class—these drafts were provided to those assigned to the *review-first* condition. While this study produced very important findings that should be followed up, replicating such a design could be difficult. Not only would one need to find instructors who would be willing to integrate an experiment as part of the classroom activities (which is difficult enough in itself), but the instructors must also be open to requiring different deadlines for each experimental condition. In the current dissertation, I designed a study that allowed me to examine the effects of providing feedback separate from the effects of receiving feedback—doing so completely behind the scenes so that the classroom assignments need not be modified.

Next, two methods—each with their own limitations—were typically used when examining the effects of peer feedback on the revision process. First, researchers might use a more qualitative approach to describe the revision process observed in small datasets. For example, Faigley and Witte (1981) tested their taxonomy of revision changes by coding and comparing only six participants of each writer level (i.e., inexperienced, advanced, and expert writers). While their taxonomy was instrumental to the understanding of the revision process was critical, their initial experiment lacked power to draw strong conclusions. A second commonly used method was to focus on the probability of feedback implementation while ignoring the quality of the actual revisions. For example, Nelson and Schunn (2009) examined how different

feedback features influenced revisions. In their study, the main dependent measure was the implementation rate (i.e., how often did the student try to address a particular comment by revising the corresponding text). While their study provided some evidence about which features of peer feedback appeared to be most helpful, it was still unclear whether these features actually led to increases in draft quality. In the current dissertation, the quality of each revision was explicitly measured in addition to the implementation rate.

Finally, prior research primarily focused on the effects of peer review on performance (i.e., a student's ability to improve a single document). For example, Tseng and Tsai (2007) had students complete multiple rounds of peer review to understand the effects of different types of peer feedback; however, the students in this study continued to revise the same document. While the findings of this study contributed to the understanding of the effects of peer feedback, it was still unclear whether the improvements students on their papers made would transfer to future writing. When the effect of peer review on future writing tasks was examined, the researchers were limited to broad conclusions rather than understanding the effects of specific components of the peer review process. For example, Lundstrom and Baker (2009) had students complete four peer review training sessions over the course of a semester. In these training sessions, students either practiced revising texts using feedback or practiced providing effective feedback. By comparing pre and post writing measures, they were able to determine whether students' writing ability improved more from only receiving feedback or only providing feedback. However, they were not able to offer any clear explanations about the causal role of particular peer reviewing elements based on their design. In the current dissertation, I was able to analyze multiple layers of the revision process as well as observe how the layers worked in combination to offer causal explanations about the revision process. In addition to completing two drafts of

the same paper, the participants also wrote a second paper on a new topic, which allowed me to observe effects of transfer.

In the current dissertation, a single experiment was conducted to examine how students of varying abilities learn from the peer review process. The writing ability of the students, the quality of the text they reviewed (i.e., the writing ability of the peer who wrote the text to be reviewed), and the ability of the feedback source (i.e., the writing ability of the peer who provided feedback) were independently manipulated. Several different measures (including draft quality, the amount and types of comments received and provided, the implementability of the comments received, and the quality of the revisions) were analyzed to answer two specific questions. In the first study, I examined how the student's writing ability and the writing ability of the feedback source affected the revision process (i.e., how the amount and types of peer feedback received, the implementability of the peer feedback, and the quality of the revisions). In the second study, I examined how the student's writing ability and the quality of the text to be reviewed affected how much students improve their writing ability from the various activities involved in reviewing peers' texts (i.e., reading the text, evaluating the quality of the paper, and detecting and diagnosing problems).

2.0 STUDY 1: EFFECTS OF RECEIVING PEER FEEDBACK ON REVISION

2.1. INTRODUCTION

In higher education, many students have few good opportunities to learn how to write better because their large content classes (i.e., more than 75 students) make it very difficult for instructors to provide adequate feedback on writing (Anrum & Roksa, 2010). One possible solution that has been gaining popularity is the use of peer review to evaluate students' writing. Peer review can benefit students in at least three general ways. First, by adding writing activities through leveraging peer review, students can experience more writing assignments, which will allow them to practice the skill of writing. Second, relative to instructor assessment alone, students can receive a larger amount of feedback from peers, which can support them in revising drafts more successfully (Cho & Schunn, 2007). Third, students can practice evaluation, detection, and diagnosis skills by providing feedback to their peers, which can improve the students' own writing ability (Wooley, et al., 2008).

Despite these benefits, instructors are reluctant to use peer review. From years of work on web-based peer review, many instructors have told me that they are unsure whether all students are capable of helping their peers. In particular, feedback from low-ability students is considered most problematic: instructors question whether low-ability students have something to offer high-ability students, and whether low-ability students helping other low-ability students would be too much like the blind leading the blind. In terms of skills, instructors worry whether low-ability students are capable of detecting problems in the text, whether they understand which problems have the largest negative impact, and whether they are able to accurately describe the

problem and offer useful solutions that would fix the major problems. Certainly the research on writing development suggests that many students struggle with those elements of writing (Flower, Hayes, Carey, Schriver, & Stratman, 1986; Hayes, Flower, Schriver, Stratman, & Carey, 1987). The broader developmental research suggests that learners can accurately detect and evaluate performance in others before they can reliably perform on their own (e.g., Crowley & Siegler, 1993), but those patterns have not been examined in complex skills like writing. Further, the literature on peer-assisted learning has some mixed results regarding relative peer ability, but those situations may differ from peer review in several important ways. Therefore, the current study addressed two research questions. The first research question focused on a pragmatically important issue—that is, how are the benefits of peer review moderated by the ability of the students (as writers and as sources of feedback)? The second research question built on peer-assisted learning theory—that is, how similar does peer feedback in writing work in comparison to other types of peer-assisted learning (e.g., peer tutoring and cooperative learning)?

Rigorous empirical work on peer review is limited, and it tends to focus on the overall effectiveness or reliability and validity (for a review, see Topping & Ehly, 1998). Effects of student ability on peer review specifically in writing have rarely been examined in prior work. In a typical setting, students are just randomly assigned to the peer review groups. Gouli, Gogoulou, and Grigoriadou (2008) described several web-based environments that support peer review, but their environment (PECASSE) was the first to offer alternative strategies to assign students to peer review groups (e.g., randomly by the system, manually by the instructor, or systematically by the system based on learners' characteristics—grouping students with similar characteristics or dissimilar characteristics). While a mechanism was offered to create peer review groups of students with the same abilities or different abilities, the authors did not provide any empirical

evidence or advice regarding which option was optimal. Crespo-Garcia and Pardo (2010) developed a system for peer review of computer programs that groups peers by similar ability levels on the assumption that being matched by ability would be best.

Only one study to date has examined how feedback from peers of similar ability or from peers of differing ability affected the peer review process. Using archival data, Patchan, Hawk, Stevens, and Schunn (under review) determined student ability level based on self-reported SAT verbal scores and looked for differences in the amount of feedback, feedback content, and implementation rates as a function of the interaction between high writing ability and low writing ability as authors and reviewers. Three patterns of results were found. First, high-ability sources provided more feedback (which was more likely to be implemented) than low-ability sources. Second, low writers received more feedback and implemented more feedback than the high writers. Most interesting were the interactions, which appeared to drive the main effects. Overall, a common pattern was found to extend across multiple layers of analyses of the comments and implementation. High writers benefited equally from low-ability sources and high-ability sources, but the low writers benefited more from the high-ability sources than the low-ability sources. The low writers received more criticism in the form of problems that focused on low prose and substance issues from high-ability sources, which led to more feedback from the high-ability sources being implemented; specifically, more comments that focused on low prose and substance were implemented. All of these benefits for low writers were large effect sizes.

The goal of the current study was to replicate the findings of Patchan, et al. (under review) in an experimental setting. This goal is important for several reasons. First, while there was random assignment of reviewer/author pairings within the Patchan et al. study, the group of peers reviewing each paper generally included both high-ability and low-ability sources. This

mixture of review sets was not a problem for describing the types of feedback that were received, but it prevented tracking of the overall effects of the feedback sources into revisions. In addition, a different course and a different writing genre will be used in the current study, which helps extend the generalizability of the results.

The next section explores the theoretical issues involved in the potential interaction between authors' writing ability and reviewers' writing ability in peer review. Given the lack of prior literature on grouping strategies specifically in writing, literature from other types of peer-assisted learning will be considered in order to develop hypotheses about how different abilities affect the benefits of peer review. Topping (2005) reviewed 25 years of development in the peer-assisted learning literature. He identified two types of peer-assisted learning that have been long established and rigorously researched: peer tutoring and cooperative learning. To map predictions, the similarities and differences between peer review (our focus) and peer tutoring and cooperative learning (existing literature) must first be unpacked. Two key dimensions will be considered: the type of interaction between students and the number of feedback ideas received. In the following section, comparisons between peer review and peer tutoring/cooperative learning on these two dimensions will be made to better understand which structural features of peer review are similar to peer tutoring and cooperative learning.

2.1.1. Comparisons between Peer Review, Peer Tutoring, and Cooperative Learning

One salient difference between peer review and peer tutoring/cooperative learning is the type of interaction between the students: bidirectional vs. one-way interaction. Within peer review, the interaction typically is one-way; especially in the cases where web environments are used to facilitate the process—that is, a reviewer submits feedback by a certain deadline and the writer

receives the feedback after the deadline has passed without any additional interaction between the students. By contrast, in peer tutoring and cooperative learning, pairs or groups of students typically work bi-directionally either face-to-face or via chat programs. Through bidirectional interaction, communication barriers between feedback provider and feedback receiver can be reduced. The greater the initial differences between provider and receiver (e.g., language, culture, ability), the more that this communication issue is likely to be important.

Another way to compare peer review, peer tutoring, and cooperative learning is to consider how much feedback/instruction is provided and received. Based on these dimensions, peer review seems to be more similar to cooperative learning. Frequently, peer review involves providing feedback to and receiving feedback from multiple peers who notice many different problems and provide many different solutions. This multiplicity is one of several advantages of peer review over instructor assessment. Several researchers have found that students who received feedback from multiple peers improved just as much as those who received feedback from a single instructor (Topping & Ehly, 1998), and in some cases, students improved more from peer feedback (Cho & Schunn, 2007). Similar to peer review, students in cooperative learning situations find themselves working with several students who each notice different things. During group work, students may need to provide explanations to several peers who have less understanding of the topic to be learned. In return, these students may receive explanations from several peers to help themselves gain a better understanding. Unlike peer review and cooperative learning, in peer tutoring, students typically interact over an extended session with only one other student, who might provide more depth on particular issues but is likely to notice fewer issues. As a result, peer tutoring may provide students with fewer learning opportunities from receiving feedback than peer review and cooperative learning.

2.1.2. Effects of Ability on Cooperative Learning and Peer Tutoring

Cooperative learning. In theory, there are several advantages to working with peers of the same ability versus working with peers of higher or lower ability (Lou et al., 1996). When students of similar abilities work together, the students are able to work at a common pace. Without the pressure of trying to keep up or waiting for others to catch up, students may be more motivated to learn. However, researchers have only speculated about this advantage and have not specifically examined whether this effect actually occurs.

By contrast, working with peers of higher or lower ability may provide opportunities to hear from multiple perspectives, which may stimulate discussion among the group members. One study that examined diversity in the workplace found that more diversity led to conflict (Jehn, Northcraft, & Neale, 1999). One thing to note is that not all conflict is negative—conflict in groups may lead to more discussion. In order to construct an outcome that will satisfy all group members, they must first engage in a discussion about each of the members' ideas. This reflection is likely to lead to a change in understanding (De Lisi & Golbeck, 1999; Moshman & Geil, 1998).

While many studies have examined the effects of group composition on cooperative learning, the majority of these studies focused on elementary or secondary students. One meta-analysis revealed that there was a small effect ($d = 0.12$) in favor of students working with peers of the same ability (Lou, et al., 1996). Interestingly, this benefit appeared to be moderated by the student's ability level—that is, the low-ability students benefited more from working with high-ability peers, and the high-ability students benefited equally from working with high-ability peers and low-ability peers. These results were based on studies from multiple contexts (e.g.,

math, science, language arts) that included wide range of ages (e.g., first grade through post-secondary school).

Relatively few studies on cooperative learning have focused on post-secondary students—none of these studies involved writing. The majority of these studies in higher education settings did not find an effect of group composition (Day et al., 2005; Goethals, 2002; Miller & Polito, 1999; Tutty & Klein, 2008; Watson & Marshall, 1995). Only two studies found an overall benefit of working with peers of the same ability rather than higher or lower ability (Cobb, 1999; Goethals, 2001). Similar to the studies focusing on younger students, several studies reported that the benefits differed for high-ability and low-ability students. Cobb (1999) and Day et al. (2005) found that only high-ability students benefited from working with other high-ability peers, and that group composition did not matter for the low-ability students. Tutty and Klein (2008) also found that high-ability students benefited from working with other high-ability peers, but the low-ability students also benefited the most from working with high-ability peers. Overall, there is not a clear pattern, except to suggest that ability level moderation does occur. The benefits of group composition would likely need to be investigated in each particular context because the pattern of results does not appear to generalize across all learning situations. Thus, an investigation in the context of university students providing feedback on writing is required.

Peer tutoring. Researchers have also addressed whether peer tutoring is more successful with peers of the same ability versus peers of higher or lower ability. This literature focuses primarily on school-age children, who either experience same-age peer tutoring (i.e., both the tutor and tutee are the same age) or cross-age tutoring (i.e., typically the tutor is older than the tutee). In a meta-analysis, Cohen, Kulik, and Kulik (1982) compared studies of same-age tutoring to cross-

age tutoring that included participants from first grade through twelfth grade in mostly math or reading. The tutees experiencing cross-age tutoring performed best. One possible benefit of cross-age tutoring is that high-ability students may be providing higher quality explanations to the low-ability students. By acting upon those explanations, the low-ability students will likely improve their understanding (Webb, Troper, & Fall, 1995). Thus, these results speak in favor of students working with peers of a different ability level.

For this age range in which students are making significant progress each year in many academic areas, students at the same age would likely be more similar in ability than students at different ages. For undergraduate students, there is likely to be a wide range of abilities and age is not likely to be an accurate predictor for general skills (Anrum & Roksa, 2010). Instead, different measures of ability should be considered.

Conclusions. Different conclusions could be drawn depending on which literature was considered. In accordance with the cooperative learning literature, high-ability students would likely benefit the most from working with other high-ability peers, but it was unclear with whom low-ability students may benefit the most. In line with the peer tutoring literature, high-ability students would likely benefit the most from working with other low-ability peers, and low-ability students would likely benefit the most from working with other high-ability peers.

2.1.3. Components of Feedback and Revision

In developing an understanding of how student ability might influence feedback in writing, it is useful to consider what kinds of feedback lead to which types of revisions with what kinds of impact on document quality. A recent study investigated which feedback features (e.g.,

summarization, identifying problems, providing solutions, localization, explanations, scope, praise, and mitigating language) were related to implementation of the feedback in the document revisions (Nelson & Schunn, 2009). Three of the features were associated with an increase in implementation: summarization (i.e., the reviewer recaps the writer's main points), localization (i.e., the reviewer identifies where in the paper the problem occurs), and solutions (i.e., the reviewer offers suggestions for how to fix a problem). If indeed these feedback features generally improve the likelihood of feedback implementation, then tracking the relative occurrence of these feedback features across writer and reviewer ability will be important.

Another recent study examined how high-ability sources and low-ability sources differed in their commenting style (Patchan, Charney, & Schunn, 2009). Interestingly, when provided with specific feedback rubrics, there were very few differences between the two types of sources. For example, both high-ability and low-ability peers produced an equal number of high-level writing comments. The one critical feature that differed was localization: high-ability sources provided more localized comments than low-ability sources. One limitation to this study was that these researchers did not examine the interactions between writer ability and feedback source. These possible interactions could be very important. For example, from a perspective of Vygotsky's (1978) zone of proximal development, low-ability sources may be less able to find problems in higher quality texts than high-ability sources. Thus, based only on the experimental data, I am limited to predicting a main effect of feedback source—that is, high-ability sources will be more likely to provide localized comments than low-ability sources, which would likely lead to higher implementation rates of these comments.

2.1.4. Competing Predictions

Previous research has provided mixed results regarding how to best assign peer review groups. Different predictions could be made depending upon which literature is considered (see Table 1). If peer review is more similar to cooperative learning, then students should benefit most from feedback provided by peers of similar ability, but this effect may be moderated by student ability. More specifically, high writers are expected to benefit from high-ability sources, but it is unclear whether low writers will benefit more from high-ability sources or low-ability sources. If peer review is more similar to peer tutoring, then students should benefit the most from feedback provided by peers of different ability. That is, high writers are expected to benefit from low-ability sources, and low writers are expected to benefit from high-ability sources.

Based on what is known from writing research, there could be two possible outcomes. First, there may be a main effect of feedback source—that is, high-ability sources are expected to provide more localized comments, which is expected to lead to higher implementation rates. Second, there may be an interaction between writer ability and feedback source—that is, high writers are expected to benefit equally from high-ability sources and low-ability sources, and low writers are expected to benefit more from high-ability sources than low-ability sources.

Table 1. Predictions of source ability for high-ability and low-ability students

Literature	High-Ability Students	Low-Ability Students
Cooperative Learning	high > low	unknown
Peer Tutoring	low > high	high > low
Writing Research[a]	high > low	high > low
Writing Research[b]	high = low	high > low

[a]based on (Patchan, et al., 2009)

[b]based on (Patchan, et al., under review)

2.2. METHOD

2.2.1. Overview

Every student provided feedback on four peers' writing using the web-based peer review functions of turnitin.com[1]. In order to examine how writer ability and feedback source affect revision activities and outcomes, first I determined the writing ability of each participant, and then I experimentally manipulated to whom the participants provided comments (consistently across all comments provided) and from whom they received comments (consistently across all comments received). In other words, participants of high-ability or low-ability provided feedback to only high-ability peers or only low-ability peers and received feedback from only high-ability

[1] The turnitin.com peer review functions primarily focused on generating end comments rather than marginalia. Reviewers were able to tag specific locations in the text that could be used in the end comment to indicate where a particular problem existed; however, this function was not obvious and most students did not use it. In addition, the specific commenting prompts were separate from the ratings prompts, which could allow one to create a reviewing assignment that utilized more fine-grained evaluation dimensions and broader commenting dimensions. Finally, the reviews were anonymous—that is, a pseudonym was used to identify both the writer and the reviewer.

peers or only low-ability peers. To examine the overall effect that receiving feedback has on revision, the quality of the participants' drafts were compared across the conditions. Most importantly, to examine how writer ability and feedback source affected the revision process, the type and amount of comments received, the implementability of the comments, and the quality of the revisions made in response to the comments received were compared across the conditions.

2.2.2. Course Context

This study was conducted in an Introduction of Psychological Science course at a very large, public research university in the US. The context was specifically selected to represent an authentic writing assignment that occurred in a large, content course as part of the Writing in the Discipline (WID) program. This course was a popular general education course that students commonly took to meet one of their social science requirements. In addition, it was compulsory for not only all psychology majors, but also for a number of other majors as well (including education and nursing). Because this course was very large (838 students), three sections (each taught by a different lecturer) were offered. Students were also required to attend one of 24 different weekly lab sections. The lab sections consisted of a smaller group of students (up to 35 students in each section). Twelve graduate student teaching assistants (TAs) taught the lab sections—each TA taught two sections.

The course was designed to expose students' to the various areas of psychology while developing their critical thinking skills. The lab sections provided the students with many opportunities to think more deeply about the material covered in the lectures. Particularly relevant to this study, students completed two short writing assignments (three pages each). For the first writing assignment, they completed a first draft, reviewed four peers' papers, and

revised a final draft; for the second writing assignment, they only completed a first draft. After each writing and reviewing task, students completed short surveys that encouraged them to reflect on the process. Students received course credit for completing all assignments, including the surveys.

2.2.3. Participants

The participants in this experiment consisted of a sample of 189 students taken from the larger course. The sampling procedure involved two steps. First, students were excluded from participation in the study analyses based on two criteria: 1) students failed to complete the background survey that was used to measure their ability, and 2) students opted out of the study. To ensure that participants received feedback from only four reviewers, a multiple of 72 participants were needed (2 writer ability x 3 text quality x 3 feedback source x 4 reviewers)[2]. Therefore, of the remaining 776 students, 648 students were randomly selected for participation. Finally, the 360 students who were randomly assigned to the *mixed* condition were excluded from the current study. The second step in our sampling procedure was to include the participants who were missing the least amount of data and ensure that there were at least 20 participants in each condition (see Table 2).

Students at all levels were represented with a predominance of less advanced students (58% freshmen, 27% sophomores, 10% juniors, and 5% seniors) as well as a great variety of majors (including social sciences, natural sciences, computer science, education, engineering, and business).

[2] As part of a larger study, there was an additional level (mixed ability) for text quality and feedback source.

Table 2. Sampling Procedure

A two-step process that maximized the amount of complete data with at least 20 participants per condition[a].

	# of Participants	Running Totals
Exclusion Criteria for the Study		838
did not complete a background survey	-61	777
opted out of the study	-1	776
random selection	-128	648
assigned to a *mixed* condition	-360	288
Inclusion Criteria for the Sample		
all assignments (3 drafts and 4 surveys)	82	82
all assignments and 3 surveys	21	103
first assignment (2 drafts) and all surveys	22	125
first assignment (2 drafts) and 3 surveys	34	159
all assignments and 2 surveys	1	160
all assignments[b]	29	189

[a]Preliminary analyses did not reveal any selection biases.

[b]Any participant who completed all of the drafts but was not yet chosen for the sample was included as additional data.

The participants' writing ability was determined by a composite of four self-reported ability measures—the average z-scores (student's score minus group mean divided by group standard deviation) of SAT verbal, SAT writing, the final grade in the first semester composition course, and the final grade in the second semester composition course (see Table 3). By using this combination of measures, I had a more generalizable ability measure that one can obtain fairly easily for use in the classroom. In addition, using a combination of multiple writing measures allowed me to consider both obvious measures that were not normed across institutions (i.e., composition grades) as well as obvious measure that were normed (i.e., SAT scores). Finally, the SAT verbal score, though not directly a measure of writing ability at first blush, was

an important contribution because it measures a variety of language related knowledge and skills that were likely to influence a writer's ability to develop fluid texts efficiently.

A median split was used to determine which students had higher writing ability (n = 93; M = .64; SD = .49) and which students had lower writing ability (n = 96; M = -0.70; SD = .47). The terms high writer and low writer will be used in this paper to indicate students with higher writing ability and students with lower writing ability. There were grouping differences of 2.8 standard deviations (i.e., a very large effect size)[3]. There were also large group differences on all components of this combined measure (see Table 3). By using a median split, a high bar was set. Likely, there were middling cases among the higher writing ability students and the lower writing ability students that could have washed out observable effects between truly high writing ability students and low writing ability students. However, using a median split was important because the observed effects will be more generalizable for classroom application.

One may wonder whether this writing ability measure is confounded with self-efficacy or motivational differences. Students completed the writing skills scale from Jones' (2008) writing self-efficacy measures, as well as, the mastery goal orientation (revised) scale and the performance-approach goal orientation (revised) scale from Midgley and colleagues' (2000) patterns of adaptive learning scales. High writers (M = 4.5, SD = 0.7) reported higher writing efficacy than low writers (M = 4.2, SD = 0.7), $t(187) = 2.7$, $p = .01$. However, high writers and low writers did not report significantly different levels of performance goals nor learning goals, $t(187) = -0.48$, $p = .63$ and $t(187) = 1.0$, $p = .32$, respectively.

[3] To further validate the composite ability measure, a 2 x 2 x 2 between subjects ANOVA revealed a significant main effect of writer ability—the high writers (M = 2.4, SD = 0.5) produced higher quality first drafts than the low writers (M = 2.0, SD = 0.4), $F(1, 181) = 22.7$, $p = .00$, $d = 0.5$.

Table 3: Summary of demographic and ability data

	High-Ability		Low-Ability	
	n	*M* (*SD*)	*n*	*M* (*SD*)
Demographics				
Gender (% female)	93	75%	95	78%
Year in School (% freshman + sophomore)	92	85%	94	84%
Age	93	18.8 (1.5)	96	19.0 (1.8)
Ability Measures				
SAT verbal	79	599 (55)	70	489 (55)
SAT writing	76	592 (63)	69	494 (58)
1st semester grade[a]	70	4.4 (0.7)	65	3.3 (0.6)
2nd semester grade[a]	49	4.2 (0.7)	51	3.1 (0.6)

[a]Composition grades were coded on a 5-point scale: 5 – placed out; 4 – A, 3 – B, 2 – C, 1 – D or below. Missing data points included participants who did not take the composition course because it was not a required course ($n_{2nd\ semester}$ = 1) and participants who were currently taking the course or will take it in the future ($n_{1st\ semester}$ = 54; $n_{2nd\ semester}$ = 88).

2.2.4. Design & Procedure

A 2 x 2 x 2 between-subjects design was used in this study (see Figure 1). The three independent variables were writer ability (high writer versus low writer), text quality reviewed[4] (high quality text versus low quality text), and feedback source (high-ability source versus low-ability source). This design allowed me to tease apart benefits of receiving feedback from benefits of providing feedback, as well as determine whether providing feedback interacted with receiving feedback, which was the focus of this study. Participants completed four main tasks: 1) write a first draft of the first assignment, 2) review peers' texts, 3) revise own text based on peer feedback, and 4) write a first draft of the second assignment. As part of a larger study, participants also completed

[4] Initial analyses included the quality of the text reviewed, but there were no significant main effects or interactions with this variable. For simplicity, the text quality variable will not be examined in this study.

a short survey after each of the main tasks; however these questions were not relevant to the current study.

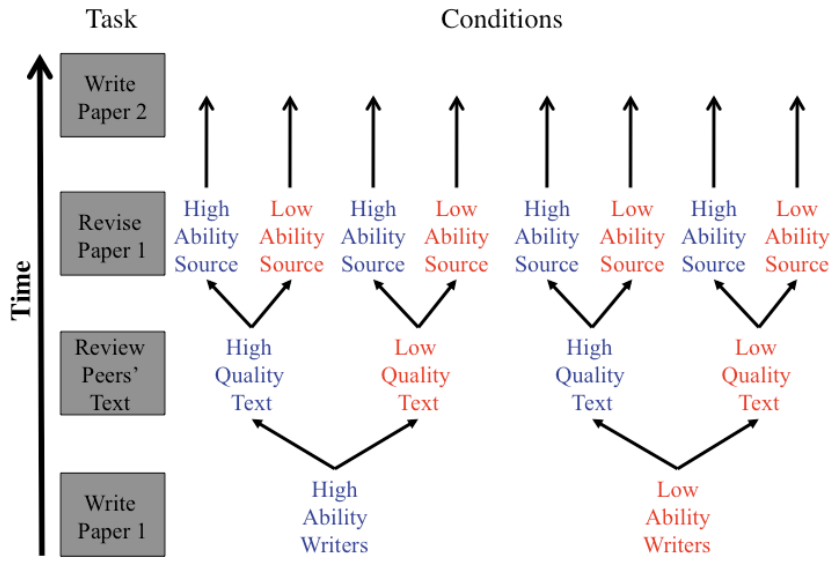

Figure 1. Between-subjects experimental design

At the end of the first month of the semester, participants had one week to write their first draft and submit it online. For this task, they were expected to write a three-page paper in which they evaluated whether MSNBC accurately reported a psychological study—applying concepts from the Research Methods chapter covered in lecture and lab in the prior week. The participants were divided into high writers and low writers based on their writing ability as described in the participants section. Comparing the quality of these texts across high writers and low writers validated the ability measure (see footnote[3]).

After the first draft deadline passed, participants were assigned four papers to review based on the text quality condition they were assigned—that is, they reviewed either four high quality texts or four low quality texts. The reviewing task involved two steps: rating the quality of the text and providing comments. The participants had one week to complete the reviews and submit them online.

Participants were able to access the peer feedback online once the reviewing deadline had passed. The source of this feedback was also manipulated: participants received feedback from peers with high writing ability (high-ability source) or peers with low writing ability (low-ability source). The participants were given one week to revise their draft based on the peer feedback. This time point was most critical for the current paper's analyses. By comparing the revised draft quality across the conditions, I could observe whether the benefit of receiving feedback differed depending on the writer ability and the feedback source. To explain how the benefit for draft quality may differ, I also analyzed the amount and type of comments received, the implementability of the comments, and the quality of the revisions that were made in response to specific comments (see Coding Process section for how these variables were measured).

Five weeks after completing the first writing assignment, the participants wrote a second paper. Again, they had one week to write a three-page paper where they evaluated how useful personality factors could be for online dating sites. The assignment details were provided two weeks after the personality chapter was covered. The purpose of this task was to see whether any of the effects found within the first assignment would transfer to future writing. In addition, possible sleeper effects could be observed—that is, reviewing peers' texts may not have an effect on the revision task, but instead, the benefits may be seen later on future writing assignments.

The TAs and lecturers were available to answer questions and offer feedback to students if more help was requested. However, most students did not take advantage of this opportunity. The TAs also provided final grades for both papers.

2.2.5. Review Support Structures

Participants were provided with a detailed rubric to use for the reviewing task. The rubric included general suggestions (e.g., be nice, be constructive, be specific) and specific guidelines, which described the three reviewing dimensions that have been applied in many disciplinary writing settings: Flow, Argument Logic, and Insight. For each commenting dimension, a number of questions were provided to prompt the reviewer to consider the paper using several particular lenses (see Appendix A). First, the flow dimension focused on whether the main ideas and the transitions between the ideas were clear. Next, the argument logic dimension focused on whether the main ideas were appropriately supported and whether obvious counter-arguments were considered. Finally, the insight dimension focused on whether a perspective beyond the assigned texts and other course materials was provided.

The purpose of these specific guidelines was to direct the participants' attention primarily towards more global writing issues (Wallace & Hayes, 1991). While different reviewing prompts would likely have produced different feedback content overall, the current study was most interested in explaining how the revision process could be affected by writer ability and feedback source, separate from the additional effects of the reviewing prompts per se. However, it is possible that detailed scaffolds of this sort for reviewing reduce differences by review source.

Finally, participants rated the quality of the documents using a 5-point scale (1–'Very Poor' to 5–'Very Good'). Within the flow dimension, two aspects of the paper were considered:

how well the paper stayed on topic and how well the paper was organized. Three aspects of the paper related to argument logic were considered: how persuasively the paper made its case, how well the author explained why causal conclusions cannot be made from correlational studies, and whether all the relevant information from the research article was provided (e.g., identified the hypothesis, the variables and their levels, the general research design, and the outcome of the study). Lastly, as a measure of insight, the participants rated how interesting and original the paper's conclusion was.

2.2.6. Coding Process

Quality of writing. Expert ratings of quality—rather than the ratings provided by the peers—were used because there could be a bias in the peer ratings as a result of the manipulations in this study. Two outside writing experts (i.e., rhetoric graduate students with extensive writing teaching experience) rated all completed drafts. Based on pilot analyses, the rubric used by the participants was revised to increase the reliability of the expert coders (see Appendix B for rubric details). While a similar 5-point scale (1–'Very Poor' to 5–'Very Good') was used, the definitions for each of the ratings were more clearly defined. In addition, some of the rating dimensions were further divided. For the flow dimension, three aspects of the paper were considered: the paragraph development, the transitions between the paragraphs, and the organization around a main idea. The aspects considered for the argument logic dimension varied slightly depending on the writing assignment. For the first writing assignment, the writing experts rated the evaluation of the article, the explanation about the casual conclusion, the description of the alternative possibilities, and the accuracy of the required information that was included. For the second writing assignment, the writing experts evaluated whether appropriate

evidence was provided and whether the authors' conclusions were relevant. Finally, for the insight dimension, they rated how well the main point was connected to a larger issue. The final inter-rater reliability was high (ICC = 0.84; Shrout & Fleiss, 1979).

Other standardized measures of draft quality (i.e., rubrics for SAT or GRE writing sections) could have been considered for this study. However, this particular measure of draft quality was chosen because it matched the goals of the course and the writing assignment. Feedback and the intended learning were focused on these dimensions, and therefore, using an assignment-aligned measure would likely strengthen the external validity of the findings to settings in which different rubrics were used.

Coding feedback. The feedback received by the participants was coded at a detailed level to determine how the amount and type of comments (i.e., a likely mediating variable for effects on document-level effects) varied as a function of writer ability and feedback source. The coding scheme originally established by Nelson and Schunn (2009) was used to categorize the types of comments, with minor revisions about how the *type of feedback* was coded (i.e., summary, praise, problem, and solution were considered independent features rather than mutually exclusive). All of the participants' comments were compiled (N = 781 reviews X 3 comment dimensions). Before categorizing the feedback, they were further segmented by idea unit because reviewers frequently commented about multiple issues within one dimension (e.g., transitions, use of examples, word choice). Therefore, a total of 7,641 received comments were coded. All undergraduate research assistants (RAs) were trained on a subset of comments, and after a moderate reliability level (Kappa = .70) was met, pairs of RAs coded all of the comments, with different pairs focusing on different coding dimensions—final Kappa values for exhaustive

coding are presented for each dimension. For comments in which the RAs disagreed, a discussion was held to determine which code would be most appropriate.

Each comment was coded for the presence/absence of three independent features: problems, solutions, and localization (Kappa = .88, .92, .63, respectively—percent agreement was always greater than 90%). A problem was considered to be present if the comment explicitly described what was wrong with the paper (e.g., "The writer did not offer insight into causal and correlational relationships."). A solution was considered to be present if the comment explicitly described how to fix a problem or improve the quality of the paper. For example, one reviewer offered the following solution: "Also, I would suggest writing a stronger conclusion to the end of the paper." A comment was considered localized when the feedback source describes where the issue occurred. Localization could take several different forms—that is, the reviewer might include a quote from the text, the reviewer might state the page and/or paragraph number, the reviewer might paraphrase the point of the text, the reviewer might mention the topic of the paragraph (e.g., "in the summary of the Weaver article"), or the reviewer might mention a general section of the paper (e.g., "in the conclusion").[5]

In addition, the focus of each comment (i.e., low prose, high prose, or substance) was coded (Kappa = .54—percent agreement was 78%). A low prose comment addressed an issue dealing with the literal text choice—usually at a word level (e.g., "Where you say 'the hypotheses and whether those hypotheses were proven', I think you would say 'that hypothesis' or 'the hypothesis' because it's just one hypothesis."). A high prose comment addressed high-level writing issues, such as the clarity of a main point or the paper in general, use of transitions, strength of arguments, provision of support and counter-arguments, inclusion of insight,

[5] Types of localization from most common to least common: text (42%), general section (21%), topic (17%), quote (12%), and paraphrase (7.8%).

introductions, and conclusions, and elaboration on vague content. For example, one peer stated, "I do not understand what the argument is as it isn't very clear." Another peer suggested, "Use your own voice in order to capture trhe [sic.] readers attention.". Finally, a substance comment addressed an issue with missing, incorrect, or contradictory content; specifically, details from the study to be summarized should have accurately included the hypothesis, the variables, the study design, the results, why causal conclusions cannot be made from correlational data, and alternative explanations to the reported results. For example, one peer pointed out, "I don't see where you stated the independent and dependent variables." Another peer recommended, "Watch your causal conclusion when you are talking about women video gamers.". Many issues can involve both high prose and substance; these comments were coded as substance.

Implementability. To examine the implementability of the comments, the same two writing experts coded whether the writer implemented a revision that addressed the issue identified in each given comment (Kappa = .74—percent agreement was 89%). Only comments that were implementable (i.e., involving a problem or solution) were coded for implementation. Microsoft Word's "compare documents" function applied to the 2nd vs. 1st drafts was used to facilitate this process. As long as the writer appeared to attempt a revision based on the comment, it was coded as implemented. Similar to coding the comments, each comment was double-coded for implementation, and all disagreements were resolved through discussion. A small percentage of comments (5%) were excluded from analysis for being too vague or unclear to determine whether they were implemented (e.g., "If anything, the paper should be a little spruced up"). While many revisions made by the writers involved mechanical or micro-level revisions, macro-level revision were coded as well (see Table 4 for examples).

Revision quality. For comments that were implemented, the same two writing experts rated the quality of the revision that was associated with the implementation, with high inter-rater reliability (ICC = .77). This quality rating was on a scale of -1 to 1[6] (see Table 4 for examples). A rating of -1 indicated a decrease in the quality of the paper. For example, one writer responded to a comment ("Consider revising your introduction. The opening quote used, in my opinion shouldn't be the first sentence") by removing the quote. In doing so, the references to "the experiments" and "this hypothesis" lacked antecedents, which made the paper less clear. A rating of 0 indicated no change in the quality of the paper. Another writer responded to a comment ("I would try and revise the last sentence in the last paragraph so that it makes more sense and you can get your point across.") by adding an additional sentence to the end of the paragraph. However, this sentence did not help make the previous sentence make more sense. Finally, a rating of 1 indicated an increase in the quality of the paper. For instance, a writer responded to a comment ("in the first sentence of the third to last paragraph: 'correlational studies cannot end in causal conclusions because the relationship between the two variables is always presently unknown.' what does 'the relationship between the two variables is always presently unknown' mean?") by adding a lengthy elaboration and clarifying the text located after the confusing prose. These revisions demonstrated the writer's understanding of the relationship between causal and correlational studies, which resulted in an increase of the quality rating of the second draft in the explanation about the casual conclusion dimension.

[6] Additional revision quality analyses were conducted that used different versions of the revision quality measure—help versus not help (which combined no difference and decrease in quality) and hurt versus not hurt (which combined no difference and increase in quality). For the help measure, the results mirrored the reported results. For the hurt measure, all results were not statistically significant.

Table 4: Examples of revisions

Rating	Comment	Draft 1 text	Draft 2 text
-1	Consider revising your introduction The opening quote used, in my opinion shouldn't be the first sentence	According to the original report, the researchers hypothesize that "adults who play video games, compared to nonplayers, would evidence poorer perceptions of their health, greater reliance on Internet-facilitated social support, more extensive media use, and a higher BMI" (Weaver 1). Throughout the studies the whole purpose was to figure out if there was a correlation between the amount of time a person played to the increase on their body weight. This hypothesis was introduced because experimenters believed that video games tend to lead to violent behaviors, obesity and muscular problems.	The whole purpose of conducting the experiments was to figure out if there was a correlation between the amounts of time a person played to the increase on their body weight. This hypothesis was introduced because scientists believed that video games tend to lead to violent behaviors, obesity and muscular problems.
0	I would try and revise the last sentence in the last paragraph so that it makes more sense and you can get your point across.	Playing video games causes bad health because there is not any physical activity evolved. Gaming only consists of sitting and pressing buttons. It can cause weight gain and it probably affects the eyes. From the lack of movement and the junk food that is most likely involved, adults will gain weight easier then juveniles due to the fact that their metabolism has slowed down since their teenage years. The fact that video gaming involves continuously looking at a screen, it most likely causes strain to the eyes.	Playing video games causes bad health because there is not any physical activity involved. Although there are a few types of interactive gaming devices, typical gaming only consists of sitting and pressing buttons. It can cause weight gain and it probably affects the eyes. From the lack of movement and the junk food that is most likely involved, adults will gain weight easier then juveniles due to the fact that their metabolism has slowed down since their teenage years. In turn, this study does not actually have any one correct answer to it because of the fact that casual conclusions cannot be formed from correlative studies. The fact that video gaming involves continuously looking at a screen, it most likely causes strain to the eyes. All I all, the overall outcome of gaming has more on a negative effect on the human body than a positive effect.
1	in the first sentence of the third to last paragraph: 'correlational studies cannot end in causal conclusions because the relationship between the two variables is always presently unknown.' what does 'the relationship between the two variables is always presently unknown' mean?	Correlational studies cannot end in causal conclusions because the relationship between the two variables is always presently unknown. The variables being studied may have the group studied in common, but that cannot prove whether one causes the other or even whether they relate at all. Correlated variables usually occur alongside each other in time, not one after another, thus there is no scientifically valid cause and effect.	Correlational studies cannot end in causal conclusions because the relationship between the two variables is always presently unknown. At any one period in time, variables with a correlational relationship can exist simultaneously. For instance, a gamer can be both depressed and fat before, during, or even after playing video games. In a correlational study, the variables are not dependent upon each other's existence in time, much unlike variables with casual natures. The variables being studied may have the 'group studied' in common, but that cannot prove whether one causes the other or even whether they relate at all. Correlated variables are capable of occurring alongside each other in time, and are not limited to one after another, thus there is no scientifically valid cause and effect on that basis alone.

[a]Relevant text was underlined.

2.3. RESULTS & DISCUSSION

2.3.1. Overview

The goal of this study was to understand how students' writer ability (as author or reviewer) influenced the benefits of receiving feedback. Specifically, I wanted to know how the writer's ability and the ability of the feedback source affected the writer's revision process, following the process from feedback received to revisions implemented. Therefore, I analyzed three aspects of the revision process: 1) the amount and type of comments received, 2) the implementability of the types of comments, and 3) the quality of the revisions that occurred when responding to a particular comment. While many analyses were conducted, only the post-hoc comparisons that focused on how the effects differed by writer ability and feedback source were discussed in the text. All descriptive and inferential statistics were reported in Appendix C and Appendix D respectively. As an indicator of effect size, Cohen's *d* (i.e., mean difference divided by average standard deviation) is also included—typically, a Cohen's *d* of .3 is considered small, .5 medium, and .8 large (J. Cohen, 1977).

Because I examine interactions of independent variables across many dependent variables, I begin with a preview of the main results. Feedback source, overall, had a number of interesting effects on the revision process, and these effects differed for high writers and low writers. In general, all writers received more comments from high-ability sources than low-ability sources, but which the specific features and types of focus were provided significantly more often by high-ability sources varied for high writers and low writers. In addition, the

impact of the feedback source on the implementability of the comments also differed for high writers and low writers—that is, few features and types of focus were influenced by the feedback source for high writers, but almost all of the features and types of focus were affected by the feedback source for low writers. While a number of significant differences between feedback sources on amount and type of feedback were found, ultimately there were no significant differences in second draft quality across the conditions. This surprising finding can be explained by two factors: 1) only a few of the features and types of focus seemed to impact the quality of the paper, and 2) feedback source did not always affect the amount and implementability of these features and types of focus—thus, canceling out an possible condition effects.

2.3.2. Revision Quality: Which Features and Focus of Comments Mattered?

First, I wanted to determine whether certain features (i.e., problem, solution, and localization) and focus (i.e., high prose or substance) of comments had more influence on the quality of the paper than did other features. Therefore, average revision quality when the feature or focus was present was compared to the average revision quality when the feature or focus was absent (i.e., did that feature or focus seem to lead to useful revisions?). As a reminder, high prose comments focused on high-level writing issues (e.g., "The very last sentence also seem a little out of place. Describing the article then saying people need to approach the media with a critical mind is a far jump. A sentence linking to two would be much easier for the reader."), and substance comments focused on the inclusion of accurate content in the paper (e.g., "One more of the counter arguments that could be added to the paper would be that people that already are depressed or have physical limitations could play video games to keep their minds off of things.").

Because the value of comments may have varied by source or recipient, the analysis was done separately for the four combinations of high writers and low writers cross with high-ability sources or low-ability sources. These analyses involved a customized ANOVA model on revision quality with main effects of writer ability, feedback source, problem, solution, localization, high prose, and substance, all two-way interactions with writer ability and feedback source, and all three-way interactions with writer ability and feedback source.

For both high writers and low writers, the explicit presence of problems or solutions did not matter—the quality of their revisions when implementing a comment with a problem or solution did not differ from the quality of their revisions when implementing a comment that did not have a problem or did not have a solution. Apparently have either explicit problem or explicit solution across all problem types was generally enough. By contrast, localization, high prose comments, and substance comments did seem to influence the revision quality. These effects differed depending on the writer ability and feedback source, so they will be described separately for each interaction case, beginning with high writers overall.

For high writers, localization and high prose comments influenced the quality of revisions, but only if the comments were provided by high-ability sources. In that case, localized comments ($M = 0.16$, $SD = 0.49$) led to poorer quality revisions than comments that were not localized ($M = 0.36$, $SD = 0.45$), $t(232) = -2.9$, $p < .01$, $d = -0.42$. This somewhat surprising result is a not easily attributable to localized comments being with a different focus because the effects of localization remain when different comment focus is taken into account; it may be that explicit localization took the thinking out of the author's hand. Higher-level focus of comments did relate to revision quality as one would expect: High prose comments ($M = 0.38$, $SD = 0.59$) led to higher quality revisions than comments that did not focus on high prose issues ($M = 0.14$,

SD = 0.51), $t(232) = 3.2$, $p < .01$, $d = 0.44$. Similar to high prose comments, substance comments (high-ability sources: *M* = 0.38, *SD* = 0.64; low-ability sources: *M* = 0.41, *SD* = 0.55) also led to higher quality revisions than comments that did not focus on substance issues (high-ability sources: *M* = 0.14, *SD* = 0.48; low-ability source: *M* = 0.12, *SD* = 0.54)—however, this effect did not differ depending on the feedback source, high-ability sources: $t(232) = 3.1$, $p < .01$, $d = 0.52$; low-ability sources: $t(164) = 3.1$, $p < .01$, $d = 0.33$.

For low writers, localization and high prose comments influenced the quality of revisions, regardless of the feedback source. Again, localized comments (high-ability sources: *M* = 0.10, *SD* = 0.54; low-ability sources: *M* = 0.20, *SD* = 0.48) were surprisingly more likely to lead to poorer quality revisions than comments that were not localized (high-ability sources: *M* = 0.26, *SD* = 0.46; low-ability sources: *M* = 0.34, *SD* = 0.45), but this time the effect held for both high-ability sources: $t(250) = -2.5$, $p = .01$, $d = -0.32$ and low-ability sources: $t(238) = -2.2$, $p = .03$, $d = -0.31$. Exactly similar as for high writers, substance comments (high-ability sources: *M* = 0.27, *SD* = 0.57; low-ability sources: *M* = 0.44, *SD* = 0.57) led to higher quality revisions than comments that did not focus on substance (high-ability sources: *M* = 0.09, *SD* = 0.47; low-ability sources: *M* = 0.10, *SD* = 0.48), high-ability sources: $t(250) = 2.4$, $p = .02$, $d = 0.33$; low-ability sources: $t(238) = 4.6$, $p < .001$, $d = 0.63$. In addition, high prose comments (*M* = 0.35, *SD* = 0.58) led to higher quality revisions than comments that did not focus on high prose issues (*M* = 0.19, *SD* = 0.49), but only different from the case of high writers because the benefit of high prose overall only held if provided by low-ability sources, $t(238) = 2.3$, $p = .03$, $d = 0.30$.

In sum, for both high writers and low writers, localization, high prose, and substance comments seemed to be more important than problems and solutions—localization had a negative impact on revision quality, while high prose and substance comments generally had a

more positive impact on revision quality than low prose comments. In addition, the interactions in these results supported the grouping of peers with similar ability, at least for low writers—that is, low writers appeared to benefit more from feedback provided by low-ability sources (specifically in the case of substance comments). For high writers, there was a mixed bag with countervailing forces at play.

2.3.3. Type and Amount of Comments: The Effects of Feedback Source

Next, I was interested in the relative amounts of comments received. In order to rule out possible confounds, the correlations between features and focus of the comments were computed (see Table 5 for proportion present and correlation matrix). There were only a few non-significant relationships—high prose issues were not at all correlated with localization and substance issues were not correlated with problems or solutions. While most of the features and focus of comments were significantly correlated with each other, all of these correlations were too small to produce confounds in analyses that examine the effects of each variable in isolation (i.e., three correlations were .1 or less and three correlations were .2 or less). Therefore, I will examine the effects of features and types of focus separately with simple correlations and ANOVAs rather than complex multiple regressions / ANCOVAs that simultaneously examine all the features.

Table 5: Proportions present for each type of comment and phi coefficients

	type of comment	N	proportion present	1	2	3
1	problem	7641	0.26	-		
2	solution	7641	0.22	.08**	-	
3	localized	7641	0.16	.16**	.20**	-
4	high prose[a]	3110	0.51	.07**	-.11**	.01
5	substance[a]	3110	0.18	-.03	0.03	-.10**

[a]high prose and substance were mutually exclusive, so their inter-correlation was not meaningful.
** $p < .01$

Given that localization, high prose, and substance seemed to have the greatest impact on the quality of the revisions (to good or bad effect), I was specifically interested in how the amount of these features and focus differed depending on writer ability and feedback source, but for completeness I also mention effects on explicit problems and solutions. A 2 x 2 between-subjects ANOVA was on the presence/absence of each feedback feature and type of focus as a dependent measure. In general, high-ability sources provided more comments than low-ability sources. However, how the relative amount of features and types of focus differed depending on the feedback source also differed depending on the writer ability. Thus, the effects of high versus low-ability source will be described separately for high writers and low writers.

For high writers, only the number of solutions and substance comments differed depending on the feedback source, and these effects were large. High writers received more comments with solutions from high-ability sources ($M = 9.4$, $SD = 4.8$) than low-ability sources ($M = 7.2$, $SD = 4.7$), $t(91) = 2.2$, $p = .03$, $d = 0.46$. In addition, high writers received more

comments that focused on substance from high-ability sources ($M = 5.6$, $SD = 3.4$) than low-ability sources ($M = 4.2$, $SD = 3.4$), $t(91) = 2.0$, $p = .05$, $d = 0.41$.

By contrast, the number of problems, localization, and substance comments that low writers received differed depending on the feedback source, again with large effects. Different from high writers, low writers received more comments with problems from high-ability sources ($M = 12.7$, $SD = 5.6$) than low-ability sources ($M = 9.2$, $SD = 5.5$), $t(94) = 3.1$, $p < .01$, $d = .63$. They also received more localized comments from high-ability sources ($M = 7.9$, $SD = 4.2$) than low-ability sources ($M = 5.9$, $SD = 4.1$), $t(94) = 2.4$, $p = .02$, $d = 0.48$. Similar to high writers, low writers receive more comments that focused on substance issues from high-ability sources ($M = 6.4$, $SD = 3.5$) than low-ability sources ($M = 4.5$, $SD = 3.4$), $t(94) = 2.7$, $p = .01$, $d = 0.55$.

Overall, writers received more feedback from high-ability sources than low-ability sources. However, the amount of specific features and types of focus differed depending on the writer ability. High writers received more solutions and substance comments from high-ability sources, and low writers received more problems, localization, and substance comments from high-ability sources. These differences between writer ability could be important. While high writers received more explicit solutions from high-ability sources, this feature did not have an influence on the quality of the revisions. Thus, this difference could be irrelevant. However, by receiving more substance comments from high-ability sources, the high writers might be more likely to improve their paper if they implemented these comments because substance comments had a greater impact on revision quality.

Similarly, low writers received more explicit problems from high-ability sources, but this feature also did not have an influence on the quality of the revisions. Consequently, this difference might not be critical in understanding the benefits of receiving feedback. More

importantly, the localized comments and substance comments were likely to impact the revisions made by low writers. While the localized comments were likely to have a negative impact on revision quality, the substance comments were likely to have a positive impact on revision quality. Despite receiving more of these two types of comments from high-ability sources, the negative and positive effects might wash out any overall benefit. Thus, low writers' text might not differ in quality based on the differences in the amount of features and types of focus provided by the different feedback sources.

2.3.4. Comment Implementability: The Effects of Feedback Source

Because I was most interested in the effects on the revision process, only comments that could be implemented (i.e., comments that described problems and/or solutions, rather than only praise or summaries) were analyzed. The proportion of implemented comments was compared across the conditions. These analyses involved a customized ANOVA model with main effects of writer ability, feedback source, problem, solution, localization, high prose, and substance, all two-way interactions with writer ability and feedback source, and all three-way interactions with writer ability and feedback source. Specifically, I wanted to determine whether the implementation rate when specific features or types of focus were present differed depending on writer ability and feedback source—in other words, were authors more or less likely to implement features from particular sources controlling for differences in type of comment provided. The effects of feedback source on implementation rates differed somewhat between high writers and low writers. Overall, writers benefited the most from receiving feedback from peers of the same ability—that is, high writers were more likely to implement comments from high-ability sources ($M = 0.37$, $SD = 0.63$) than low-ability sources source ($M = 0.28$, $SD = 0.63$), $t(1246) = 2.4$, $p =$

.02, $d = 0.14$, and low writers were more likely to implement comments from high-ability sources ($M = 0.30$, $SD = 0.63$) than low-ability sources source ($M = 0.44$, $SD = 0.65$), $t(1464) = -4.1$, $p < .001$, $d = -0.22$. The remainder of effects for high writers and low writers will be described separately.

Moving to more specific comment features and types, the source of the feedback rarely influenced the likelihood that high writers would implement feedback—that is, regardless of the feature or type of focus, comments were typically implemented equally often regardless of the source. The only significant difference in implementation rates for high writers was for substance comments, which presumably drove the overall implementation difference effect. High writers were more likely to implement substance comments when provided by high-ability sources ($M = 0.40$, $SD = 0.67$) than low-ability sources source ($M = 0.25$, $SD = 0.65$), $t(436) = 2.4$, $p = .02$, $d = 0.23$.

By contrast, the source of the feedback almost always influenced the likelihood of implementation for low writers—that is, regardless of the feature or focus, low writers were more likely to implement feedback provided by low-ability sources than high-ability sources. The only non-significant difference in implementation rates for low writers was for solutions. Low writers were only marginally more likely to implement solutions when provided by low-ability sources ($M = 0.40$, $SD = 0.58$) than high-ability sources source ($M = 0.33$, $SD = 0.53$), $t(827) = -1.8$, $p = .07$, $d = -0.13$. In other words, low-ability writers found low-ability source comments generally more implementable.

The different effects of feedback source on implementation rates would likely impact the overall benefit of receiving feedback. For high writers, the only difference between high-ability sources and low-ability sources was the implementability of substance comments. This

difference was especially important because substance comments were most likely to affect the revision quality. High writers were more likely to receive substance comments from high-ability sources, and consistently, they were also more likely to implement these comments when provided by high-ability sources.

Understanding how the effects of feedback source on implementation rates influenced the benefit of receiving feedback for low writers is more complicated. While they were more likely to implement problems when provided by low-ability sources, these revisions were not likely to affect the quality of the paper. More importantly were the differences for localized comments, high prose comments, and substance comments. First, low writers were more likely to implement localized comments when provided by low-ability sources. This result could be problematic because localized comments tended to lead to poorer revision quality. However, the number of implemented localized comments could be inconsequential because low writers were more likely to receive localized comments from high-ability sources than low-ability sources. Next, low writers were more likely to implement high prose comments from low-ability sources than high-ability sources. This result was especially important because high prose comments had a positive impact on revision quality, but only when provided by low-ability sources. Even though low-ability sources provided the same amount of high prose comments as high-ability sources, the low writers were more likely to implement these comments when provided by low-ability sources and the revisions were likely to be of higher quality. Finally, low writers were more likely to implement substance comments when provided by low-ability sources rather than high-ability sources. Unfortunately, this benefit would likely be reduced because low writers were more likely to receive substance comments from high-ability sources. In other words, for low-ability writers, there were often opposing effects between how likely they were to receive

comments from high and low sources and how likely they were to implement comments received from high and low sources.

2.3.5. Draft Quality: The Effects of Feedback Source on Revision and Future Writing

Overall, writers improved their papers between first draft and second draft, as well as between the first paper and second paper. While these findings were encouraging, the emphasis of this study was to understand the effects of feedback source on revisions and future writing—that is, did these interactions in feedback input produce differences in final draft scores of the first paper?

The countervailing effects at the amount and implementation levels would suggest no overall difference in impact. A 2 x 2 between subjects ANCOVA with first draft quality as a covariate revealed neither a significant main effect of feedback source nor a significant interaction between feedback source and text quality for high writers or low writers. High writers who received feedback from high feedback sources (M = 2.6, SD = 0.2) wrote second drafts of the same quality as high writers who received feedback from low feedback sources (M = 2.6, SD = 0.2), $t(91) < 1$, p = .40, d = 0.17. In addition, low writers who received feedback from high feedback sources (M = 2.2, SD = 0.2) wrote second drafts of the same quality as low writers who received feedback from low feedback sources (M = 2.3, SD = 0.2), $t(94) < 1$, p = .40, d = -0.17. These results indicated that receiving feedback helped all participants equally on the revision task, regardless of the feedback source. One may wonder whether similar results would be found for the specific reviewing dimensions (e.g., flow, logic). Indeed, similar results were found—all participants benefited equally regardless of the feedback source.

Similarly, there were no transfer effects, as measured by the quality of the second paper. A 2 x 2 between subjects ANCOVA on the quality of the second paper with first draft quality of the first paper as a covariate revealed neither a significant main effect of feedback source nor a significant interaction between feedback source and text quality for high writers or low writers. High writers who received feedback from high feedback sources on the first paper (M = 2.5, SD = 0.5) wrote second papers of the same quality as high writers who received feedback from low feedback sources on the first paper (M = 2.6, SD = 0.5), $t(62) < 1$, p = .62, d = -0.13. Similarly, low writers who received feedback from high feedback sources on the first paper (M = 2.3, SD = 0.4) wrote second papers of the same quality as low writers who received feedback from low feedback sources on the first paper (M = 2.2, SD = 0.4), $t(65) = 1.2$, p = .24, d = 0.29. Again, this result indicated that receiving feedback helped all participants equally on future writing, regardless of the feedback source. Likewise, all participants continued to benefit equally regardless of the feedback source for each of the specific reviewing dimensions.

2.4. GENERAL DISCUSSION

2.4.1. Summary of Results

Three features and types of focus were identified as important to the revision process—that is, localization, high prose, and substance comments influenced the paper quality more than problems and solutions. Specifically, localization had a negative impact on revision quality. In contrast, high prose comments (i.e., comments focused on high-level writing issues, such as "To me, rather than posing an argument, it just seems as though you are retelling the study. Take a

side and back it up using details from the articles and also bring in outside information that will help support your position, just make sure it is revelant the the articles.") and substance comments (i.e., comments focused on the inclusion of accurate content in the paper, such as "Explain in more depth why causal conclusions cannot be made from the correlations.") had a greater positive impact on revision quality. While these comments were consistently important for high writers and low writers, the effect of feedback source on the amount received and implementability differed between high writers and low writers.

The feedback source seemed to have only a minimal effect on high writers' revision. High writers received more substance comments from high-ability sources than low-ability sources. Furthermore, high writers were more likely to implement the substance comments from high-ability sources. These differences suggested that high writers could benefit more from receiving feedback from high-ability sources.

Similarly, low writers received more substance comments from high-ability sources than low-ability sources; however, they were only more likely to implement substance comments if provided by low-ability sources. This inconsistent effect of feedback source likely decreased the overall benefit of implementing substance comments. In addition, low writers received more localized comments from high-ability sources. While implementing localized comments appeared to hurt the quality of the paper, this negative effect was likely to be minimal because low writers were less likely to implement localized comment from high-ability sources. Although low writers did not receive high prose comments more from high-ability sources or low-ability sources, they were more likely to implement high prose comments from low-ability sources. Therefore, by focusing on the high prose comments from low-ability sources, low

writers could make a positive impact on their paper. Overall, these differences suggested that low writers could benefit more from receiving feedback from high-ability sources.

Finally, there were other effects of feedback source that were not likely to impact the overall draft quality. First, high writers received more solutions from high-ability sources. This difference would not likely impact the overall quality of the paper because comments with explicit solutions offered were not more likely to lead to high quality or lower quality revisions. Likewise, low writers received more problems from high-ability sources, and they were more likely to implement problems from low-ability sources. These differences would also not likely impact the overall quality of the paper because comments with explicit problems described were not likely to lead to higher quality or lower quality revisions. Moreover, the inconsistent effects of feedback source would further minimize any possible benefits or detriments.

2.4.2. Comparisons to Prior Literature

Based on prior research, it was unclear how to best assign peer review groups because the empirical evidence has been mixed. I considered different predictions depending upon which literature was examined. According to the cooperative learning literature, high writers should benefit more from high-ability sources, but low writers may benefit from high-ability sources and/or low-ability sources. According to the peer tutoring literature, high writers should benefit from low-ability sources, and low writers should benefit from high-ability sources. Finally, according to the writing literature, two possible expectations could occur: 1) all students should benefit more from high-ability sources, or 2) only low writers should benefit from high-ability sources, and high writers should benefit equally from high-ability sources and low-ability sources. Based on the findings of the current study, peer review was more similar to the

cooperative learning literature—high writers benefited more from high-ability sources, and low writers benefited more from low-ability sources.

2.4.3. Implications for Classroom Practice

Overall, there were relatively few significant effects of feedback source—suggesting that peer review is pretty safe from a receiving feedback perspective. There were several findings from the current study that lead to the tentative implication that writers could benefit more with feedback from of peers of a similar ability—that is, high writers appeared to benefit more from feedback provided by high-ability sources, and low writers appeared to benefit more from feedback provided by low-ability sources.

Based on these findings, the benefits of peer review might be maximized with different instructional interventions that are adapted specifically for high-ability writers and low-ability writers. Advances in intelligent systems research could help facilitate the adaptable instructional interventions. By using these new research methods, one could develop a peer-review system that is able to identify which students experience similar problems so that they can be grouped together—therefore, possibly maximizing the potential learning benefits. Finally, self-reported measures appeared to be reasonably easy to obtain surrogates for writing ability that instructors could use to create these groups.

3.0 STUDY 2: UNDERSTANDING THE BENEFITS OF REVIEWING PEERS' TEXTS

3.1. INTRODUCTION

Peer editing is nothing more than the blind leading the blind with unskilled editors guiding inexperienced writers in a process neither understands well.

Michael Graner (1987, p. 40)

A common concern of many instructors have about peer review is that some students may not receive helpful feedback. Peer review is very time consuming, and if students are not being helped with the feedback they receive, then time that could have been spent on other instructional activities might be considered wasted. Despite these concerns, empirical research has provided support for peer review as an overall activity, showing that it can be just as effective as only having an instructor review (Topping, 2005), and in some cases, it has even been found to be more effective (Cho & Schunn, 2007). One possibility is that the educative value in peer review is not primarily from receiving feedback, but instead, students may learn more about writing from being the evaluator.

This contrast between receiving and providing was powerfully illustrated in a recent study by Lundstrom and Baker (2009) of second language learners. They examined whether students' writing benefited most from only providing feedback or only receiving feedback. Students experienced four peer-review training sessions throughout the course of a semester. During the training sessions, students were given sample essays, which they either practiced how to revise the given essay based on provided canned feedback (i.e., they only revised papers using this feedback) or practiced how to provide effective feedback (i.e., they only provided feedback

and did no paper revising). By comparing pre and post writing measures, they found that lower ability writers who only provided feedback made greater writing gains than lower ability writers who only received feedback; a large difference between conditions of approximately one standard deviation. This same pattern held for intermediate level writers without prior peer review experience. It was only intermediate learners with prior peer review experience who benefited equally from providing or receiving feedback; given the second language learning context, there were no advanced learners in the study. Overall, these results suggested that providing feedback could be more important than receiving feedback, although without clear explanations for the causes of neither the overall benefits nor the interactions with learner type.

Understanding how reviewing peers' texts benefits students is an important practical question for writing instruction because of the large time demands for providing reviews to peers and because of the resistance to peer review activities by some students and instructors (Kaufman & Schunn, in press). In addition, there is the important theoretical question of how peer review contributes to student learning. Therefore, the goal of the current paper is to understand the educative value of providing feedback.

3.1.1. Learning Opportunities during Peer Review

Typically, students complete two tasks while reviewing peers' papers—first, students rate the quality of the paper based on the specified grading criteria, and then they offer the peer feedback about the major problems and how to fix them. For each of these tasks, students complete a number of subtasks—each a possible learning opportunity for the student to improve their ability to revise while writing.

The first peer review subtask a student must complete is reading. By reading peers' texts, students are exposed to several different models of writing. From these models that are likely to vary in quality, students could observe what strategies led to successful writing. For example, Charney and Carlson (1995) found that students wrote more organized method sections after receiving models than students who did not receive any models. Interestingly, students in that study benefited equally from seeing only high quality models and mixed quality models (as long as one of the models was high quality), but they benefited less from receiving only poor and moderate quality models. These results suggested that the quality models have an impact on how much could be learned from models of writing. Therefore, if students benefit from peer review via reading models, then I would expect that reviewing high quality texts would be of more value than reviewing low quality texts. Despite these benefits of models, only receiving models may not be the optimal way to learn about writing. A more recent study found that students who read three models twice wrote lower quality texts than those who peer reviewed the three models (Cho & Schunn, 2010). These results suggested that to maximize learning, the student must do more than just reading.

The next peer review subtask that the student must complete is evaluation. This subtask can take on several forms. The most obvious is the evaluation of the overall quality of the paper, which is necessary for the student to determine which quality rating is most appropriate. Another form of evaluation that students should complete is evaluating the severity of the problems detected. This form of evaluation is especially important when deciding which problems to include in the feedback. Several researchers have found that students were able to improve their writing ability from evaluating texts. Students who were trained to use a rubric and evaluate a paper showed greater improvements between the pre-test and post-test writing assignments than

students who just learned about the components of writing (Sager, 1973). Although statistical analyses were not formally reported, Graner (1987) found evidence suggesting that students who evaluated two texts were equally likely to revise their papers successfully (i.e., significant improvements between first and second drafts) as students who received feedback from peers. These results indicated that learning during evaluation could be just as effective as learning from feedback. However, similar to learning from models, evaluation alone may not be the optimal way to learn about writing. Students who only provided quality ratings on peers' papers wrote poorer quality papers than those who also provided feedback in addition to the ratings (Wooley, et al., 2008). Therefore, while reading models and evaluating text offer students possible opportunities to learn about writing, the task of composing feedback (i.e., problem detection and problem diagnosis) seems to be necessary to maximize learning benefits.

The final two peer review subtasks that the student must complete are particularly important for writing feedback. First, a student must detect problems in the peer's text. Then the student must diagnose the problem—that is, once the student has chosen which problems to include in the feedback, information about those problems (e.g., what is the problem, why is it problematic, how to fix it) must be described. Evidence from the Wooley, et al. (2008) study seems to support the need to practice detecting and diagnosing problems for students to receive greater learning benefits.

The goal of the current study was to understand whether the extent to which a student could learn from reviewing peers' texts differed depending on the quality of the text being reviewed and whether these differences were further moderated by the student's ability. This goal is important for instructors who want to design assignments that maximize learning benefits. However, of the limited prior literature that has examined the effects of various subtasks of

reviewing peers' texts, none of the researchers specifically examine how the effects differed depending on the writing ability of the students. Therefore, the current study will explore these effects.

3.2. METHOD

3.2.1. Overview

The method and data source used in the current study was the same as the method and data source used in Study 1. For completeness, the basic approach is summarized again here, and there will be only minor variations in the prose that reflect the different data elements that are important to address the main questions of Study 2.

Every student provided feedback on four peers' writing using the web-based peer review functions of turnitin.com[7]. In order to describe the extent to which providing comments benefited students, first I determined the writing ability of each participant, and then I experimentally manipulated to whom the participants provided comments (consistently across all comments provided) and from whom they received comments (consistently across all comments received). In other words, participants of high-ability or low-ability provided feedback to only high-ability peers or only low-ability peers and received feedback from only high-ability peers or

[7] The turnitin.com peer review functions primarily focused on generating end comments rather than marginalia. Reviewers were able to tag specific locations in the text that could be used in the end comment to indicate where a particular problem existed; however, this function was not obvious and most students did not use it. In addition, the specific commenting prompts were separate from the ratings prompts, which could allow one to create a reviewing assignment that utilized more fine-grained evaluation dimensions and broader commenting dimensions. Finally, the reviews were anonymous—that is, a pseudonym was used to identify both the writer and the reviewer.

only low-ability peers. To examine the overall effect that providing feedback has on revision and future writing, the quality of the participants' drafts were compared across the conditions. In addition, to examine how writer ability and text quality affected how much students learn from peer review, the amount, features, and focus of comments provided were compared across the conditions.

3.2.2. Course Context

This study was conducted in an Introduction of Psychological Science course at a very large, public research university in the US. The context was specifically selected to represent an authentic writing assignment that occurred in a large, content course as part of the Writing in the Discipline (WID) program. This course was a popular general education course that students commonly took to meet one of their social science requirements. In addition, it was compulsory for not only all psychology majors, but also for a number of other majors as well (including education and nursing). Because this course was very large (838 students), three sections (each taught by a different lecturer) were offered. Students were also required to attend one of 24 different weekly lab sections. The lab sections consisted of a smaller group of students (up to 35 students in each section). Twelve graduate student teaching assistants (TAs) taught the lab sections—each TA taught two sections.

The course was designed to expose students' to the various areas of psychology while developing their critical thinking skills. The lab sections provided the students with many opportunities to think more deeply about the material covered in the lectures. Particularly relevant to this study, students completed two short writing assignments (three pages each). For the first writing assignment, they completed a first draft, reviewed four peers' papers, and

revised a final draft; for the second writing assignment, they only completed a first draft. After each writing and reviewing task, students completed short surveys that encouraged them to reflect on the process. Students received course credit for completing all assignments, including the surveys.

3.2.3. Participants

The participants consisted of a sample of 189 students taken from the larger course. The sampling procedure involved two steps. First, students were excluded from participation in the study analyses based on two criteria: 1) students failed to complete the background survey that was used to measure their ability, and 2) students opted out of the study. To ensure that participants received feedback from only four reviewers, a multiple of 72 participants were needed (2 writer ability x 3 text quality x 3 feedback source x 4 reviewers)[8]. Therefore, of the remaining 776 students, 648 students were randomly selected for participation. Finally, the 360 students who were randomly assigned to the *mixed* condition were excluded from the current study. The second step in our sampling procedure was to include the participants who were missing the least amount of data and ensure that there were at least 20 participants in each condition (see Table 6).

Students at all levels were represented with a predominance of less advanced students (58% freshmen, 27% sophomores, 10% juniors, and 5% seniors) as well as a great variety of majors (including social sciences, natural sciences, computer science, education, engineering, and business).

[8] As part of a larger study, there was an additional level (mixed ability) for text quality and feedback source.

Table 6: Sampling Procedure

A two-step process that maximized the amount of complete data with at least 20 participants per condition[a]

	# of Participants	Running Totals
Exclusion Criteria for the Study		838
did not complete a background survey	-61	777
opted out of the study	-1	776
random selection	-128	648
assigned to a *mixed* condition	-360	288
Inclusion Criteria for the Sample		
all assignments (3 drafts and 4 surveys)	82	82
all assignments and 3 surveys	21	103
first assignment (2 drafts) and all surveys	22	125
first assignment (2 drafts) and 3 surveys	34	159
all assignments and 2 surveys	1	160
all assignments[b]	29	189

[a]Preliminary analyses did not reveal any selection biases by demographic characteristics.

[b]Any participant who completed all of the drafts but was not yet chosen for the sample was included as additional data.

The participants' writing ability was determined by a composite of four self-reported ability measures—that is, the average z-scores (student's score minus group mean divided by group standard deviation) of SAT verbal, SAT writing, the final grade in the first semester composition course, and the final grade in the second semester composition course (see Table 7). By using this combination of measures, I had a more generalizable ability measure that one can obtain fairly easily for use in the classroom. In addition, using a combination of multiple writing measures allowed me to consider both obvious measures that were not normed across institutions (i.e., composition grades) as well as obvious measure that were normed (i.e., SAT scores). Finally, the SAT verbal score, though not directly a measure of writing ability at first blush, was

an important contribution because it measures a variety of language related knowledge and skills that were likely to influence a writer's ability to develop fluid texts efficiently.

A median split was used to determine which students had high writing ability (n = 93; M = .64; SD = .49) and which students had low writing ability (n = 96; M = -0.70; SD = .47) . The terms high writer and low writer will be used in this paper to indicate students with higher writing ability and students with lower writing ability. There were grouping differences of 2.8 standard deviations (i.e., a very large effect size)[9]. There were also large group differences on all components of this combined measure (see Table 7). By using a median split, a high bar was set. Likely, there were middling cases among the high writing ability students and the low writing ability students that could have washed out observable effects between truly high writing ability students and low writing ability students. However, using a median split was important because the observed effects will be more generalizable for classroom application.

[9] To further validate the composite ability measure, a 2 x 2 x 2 between subjects ANOVA revealed a significant main effect of writer ability—the high writers (M = 2.4, SD = 0.5) produced higher quality first drafts than the low writers (M = 2.0, SD = 0.4), $F(1, 181) = 22.7$, $p = .00$, $d = 0.5$.

Table 7: Summary of demographic and ability data

	High-Ability		Low-Ability	
	n	*M* (*SD*)	*n*	*M* (*SD*)
Demographics				
Gender (% female)	93	75%	95	78%
Year in School (% freshman + sophomore)	92	85%	94	84%
Age	93	18.8 (1.5)	96	19.0 (1.8)
Ability Measures				
SAT verbal	79	599 (55)	70	489 (55)
SAT writing	76	592 (63)	69	494 (58)
1st semester grade[a]	70	4.4 (0.7)	65	3.3 (0.6)
2nd semester grade[a]	49	4.2 (0.7)	51	3.1 (0.6)

[a]Composition grades were coded on a 5-point scale: 5 – placed out; 4 – A, 3 – B, 2 – C, 1 – D or below. Missing data points included participants who did not take the composition course because it was not a required course ($n_{2nd\ semester} = 1$) and participants who were currently taking the course or will take it in the future ($n_{1st\ semester} = 54$; $n_{2nd\ semester} = 88$).

3.2.4. Design & Procedure

A 2 x 2 x 2 between-subjects design was used in this study (see Figure 2). The three independent variables were writer ability (high writer versus low writer), text quality reviewed (high quality text versus low quality text), and feedback source[10] (high ability source versus low ability source). This design allowed me to tease apart benefits of providing feedback from benefits of receiving feedback, as well as determine whether receiving feedback interacted with providing feedback, which was the focus of this study. Participants completed four main tasks: 1) write a first draft of the first assignment, 2) review peers' texts, 3) revise own text based on peer feedback, and 4) write a first draft of the second assignment. As part of a larger study,

[10] Initial analyses included the feedback source, but there were no significant main effects or interactions with this variable. For simplicity, the feedback source variable will not be examined in this study.

participants also completed a short survey after each of the main tasks; however these questions were not relevant to the current study.

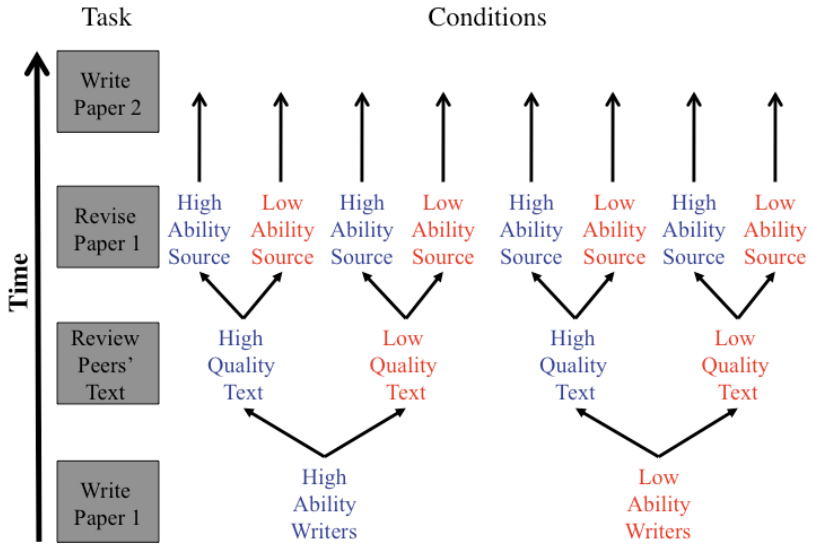

Figure 2. Between-subjects experimental design

At the end of the first month of the semester, participants had one week to write their first draft and submit it online. For this task, they were expected to write a three-page paper in which they evaluated whether MSNBC accurately reported a psychological study—applying concepts from the Research Methods chapter covered in lecture and lab in the prior week. The participants were divided into high writers and low writers based on their writing ability as described in the participants section. Comparing the quality of these texts across high writers and low writers validated the ability measure (see footnote[9]).

After the first draft deadline passed, participants were assigned four papers to review based on the text quality condition they were assigned—that is, they reviewed either four high

quality texts or four low quality texts. The reviewing task involved two steps: rating the quality of the text and providing comments. The participants had one week to complete the reviews and submit them online. This time point was most critical for the current paper's analyses. The amount, features, and focus of comments provided by the participants could be examined to see whether they get more practice detecting and diagnosing problems in low quality texts than the high quality texts (see Coding Process section for how these variables were measured).

Participants were able to access the peer feedback online once the reviewing deadline had passed. The source of this feedback was also manipulated: participants received feedback from peers with high writing ability (high ability source) or peers with low writing ability (low ability source). The participants were given one week to revise their draft based on the peer feedback.

Five weeks after completing the first writing assignment, the participants wrote a second paper. Again, they had one week to write a three-page paper where they evaluated how useful personality factors could be for online dating sites. The assignment details were provided two weeks after the personality chapter was covered. The purpose of this task was to see whether any of the effects found within the first assignment would transfer to future writing. In addition, possible sleeper effects could be observed—that is, reviewing peers' texts may not have an effect on the revision task, but instead, the benefits may be seen later on future writing assignments.

The TAs and lecturers were available to answer questions and offer feedback to students if more help was requested. However, most students did not take advantage of this opportunity. The TAs also provided final grades for both papers.

3.2.5. Review Support Structures

Participants were provided with a detailed rubric to use for the reviewing task. The rubric included general suggestions (e.g., be nice, be constructive, be specific) and specific guidelines, which described the three reviewing dimensions that have been applied in many disciplinary writing settings: Flow, Argument Logic, and Insight. For each commenting dimension, a number of questions were provided to prompt the reviewer to consider the paper using several particular lenses (see Appendix A). First, the flow dimension focused on whether the main ideas and the transitions between the ideas were clear. Next, the argument logic dimension focused on whether the main ideas were appropriately supported and whether obvious counter-arguments were considered. Finally, the insight dimension focused on whether a perspective beyond the assigned texts and other course materials was provided.

The purpose of these specific guidelines was to direct the participants' attention primarily towards global writing issues (Wallace & Hayes, 1991). While different reviewing prompts would likely have produced different feedback content overall, the current study was most interested in explaining how the benefit of providing feedback could be affected by writer ability and text quality, separate from the additional effects of the reviewing prompts per se. However, it is possible that detailed scaffolds of this sort for reviewing reduce differences by text quality. For example, when providing feedback, students may try to comment on all of the prompts regardless of the quality of the paper.

Finally, participants rated the quality of the documents using a 5-point scale (1–'Very Poor' to 5–'Very Good'). Within the flow dimension, two aspects of the paper were considered: how well the paper stayed on topic and how well the paper was organized. Three aspects of the paper related to argument logic were considered: how persuasively the paper made its case, how

well the author explained why causal conclusions cannot be made from correlational studies, and whether all the relevant information from the research article was provided (e.g., identified the hypothesis, the variables and their levels, the general research design, and the outcome of the study). Lastly, as a measure of insight, the participants rated how interesting and original the paper's conclusion was.

3.2.6. Coding Process

Quality of writing. Expert ratings of quality—rather than the ratings provided by the peers—were used because there could be a bias in the peer ratings as a result of the manipulations in this study. Two outside writing experts (i.e., rhetoric graduate students with extensive writing teaching experience) rated all completed drafts. Based on pilot analyses, the rubric used by the participants was revised to increase the reliability of the expert coders (see Appendix B for rubric details). While a similar 5-point scale (1–'Very Poor' to 5–'Very Good') was used, the definitions for each of the ratings were more clearly defined. In addition, some of the rating dimensions were further divided. For the flow dimension, three aspects of the paper were considered: the paragraph development, the transitions between the paragraphs, and the organization around a main idea. The aspects considered for the argument logic dimension varied slightly depending on the writing assignment. For the first writing assignment, the writing experts rated the evaluation of the article, the explanation about the casual conclusion, the description of the alternative possibilities, and the accuracy of the required information that was included. For the second writing assignment, the writing experts evaluated whether appropriate evidence was provided and whether the authors' conclusions were relevant. Finally, for the

insight dimension, they rated how well the main point was connected to a larger issue. The final inter-rater reliability was high (ICC = 0.84; Shrout & Fleiss, 1979).

Other standardized measures of draft quality (i.e., rubrics for SAT or GRE writing sections) could have been considered for this study. However, this particular measure of draft quality was chosen because it matched the goals of the course and the writing assignment. Feedback and the intended learning was focused on these dimensions, and therefore, using an assignment-aligned measure would likely strengthen the external validity of the findings to settings in which different rubrics were used.

Coding feedback. The feedback provided by the participants was coded at a detailed level to determine how the amount, features, and focus of comments varied as a function of writer ability and paper quality. The coding scheme originally established by Nelson and Schunn (2009) was used to categorize the types of comments, with minor revisions about how the *type of feedback* was coded (i.e., summary, praise, problem, and solution were considered independent features rather than mutually exclusive). All of the participants' comments were compiled (N = 811 reviews X 3 comment dimensions). Before categorizing the feedback, they were further segmented by idea unit because reviewers frequently commented about multiple issues within one dimension (e.g., transitions, use of examples, word choice). Therefore, a total of 8,020 provided comments were coded and analyzed. All undergraduate research assistants (RAs) were trained on a subset of comments, and after a moderate reliability level (Kappa = .70) was met, pairs of RAs coded all of the comments with different pairs focusing on different coding dimensions—final Kappa values for exhaustive coding are presented for each dimension. For comments in which the RAs disagreed, a discussion was held to determine which code would be most appropriate.

Each comment was coded for the presence/absence of three independent features: problems, solutions, and localization (Kappa = .88, .92, .63, respectively—percent agreement was always greater than 90%). A problem was considered to be present if the comment explicitly described what was wrong with the paper (e.g., "The writer did not offer insight into causal and correlational relationships."). A solution was considered to be present if the comment explicitly described how to fix a problem or improve the quality of the paper. For example, one reviewer offered the following solution: "Also, I would suggest writing a stronger conclusion to the end of the paper." A comment was considered localized when the feedback source describes where the issue occurred. Localization could take several different forms—that is, the reviewer might include a quote from the text, the reviewer might state the page and/or paragraph number, the reviewer might paraphrase the point of the text, the reviewer might mention the topic of the paragraph (e.g., "in the summary of the Weaver article"), or the reviewer might mention a general section of the paper (e.g., "in the conclusion").[11]

In addition, the focus of each comment (i.e., low prose, high prose, or substance) was coded (Kappa = .54—percent agreement was 78%). A low prose comment addressed an issue dealing with the literal text choice—usually at a word level (e.g., "Where you say 'the hypotheses and whether those hypotheses were proven', I think you would say 'that hypothesis' or 'the hypothesis' because it's just one hypothesis."). A high prose comment addressed high-level writing issues, such as the clarity of a main point or the paper in general, use of transitions, strength of arguments, provision of support and counter-arguments, inclusion of insight, introductions, and conclusions, and elaboration on vague content. For example, one peer stated, "I do not understand what the argument is as it isn't very clear." Another peer suggested, "Use

[11] Types of localization from most common to least common: text (42%), general section (21%), topic (17%), quote (12%), and paraphrase (7.8%).

your own voice in order to capture trhe [sic.] readers attention.". Finally, a substance comment addressed an issue with missing, incorrect, or contradictory content; specifically, details from the study to be summarized should have accurately included the hypothesis, the variables, the study design, the results, why causal conclusions cannot be made from correlational data, and alternative explanations to the reported results. For example, one peer pointed out, "I don't see where you stated the independent and dependent variables." Another peer recommended, "Watch your causal conclusion when you are talking about women video gamers.". Many issues can involve both high prose and substance; these comments were always coded as substance.

3.3. RESULTS & DISCUSSION

3.3.1. Overview

The goal of this study was to understand whether the extent to which a student could learn from reviewing peers' texts differed depending on the quality of the text being reviewed and whether these differences were further moderated by the student's ability. First, I examined draft quality differences across the conditions to see whether they matched any of the predicted patterns. Then, I observed how the amount, features (i.e., problem, solution, and localization), and focus (i.e., high prose or substance) of comments differed across the conditions. As a reminder, high prose comments focused on high-level writing issues (e.g., "The very last sentence also seem a little out of place. Describing the article then saying people need to approach the media with a critical mind is a far jump. A sentence linking to two would be much easier for the reader."), and substance comments focused on the inclusion of accurate content in the paper (e.g., "One more

of the counter arguments that could be added to the paper would be that people that already are depressed or have physical limitations could play video games to keep their minds off of things.").

While many analyses were conducted, only the post-hoc comparisons that focused on how the effects differed by writer ability and text quality were discussed in the text. All inferential statistics are reported in Appendix E. As an indicator of effect size, Cohen's *d* (i.e., mean difference divided by average standard deviation) is also included—typically, a Cohen's *d* of .3 is considered small, .5 is medium, and .8 is large (J. Cohen, 1977).

A 2 x 2 x 2 between subjects ANOVA revealed a marginal main effect of text quality on first draft quality; specifically this difference was localized to low writers only. Because there was no logical reason for the quality of the to-be-reviewed documents to affect the quality of the first draft, this difference must have resulted from noise in random assignment to condition or in the data. The remaining analyses adjusted for this difference by including the quality of first draft as a covariate. In addition, high-ability participants and low-ability participants were analyzed separately in order to keep the focus of this study on the effects of text quality on writer ability rather than comparing these two groups.

3.3.2. Draft Quality

Overall, writers improved their papers between first draft and second draft, as well as between the first paper and second paper. While these findings were encouraging, the emphasis of this study was to understand the effects of the quality of the text reviewed on revisions and future writing.

Condition effects on revision. There were no significant main effects or interactions on the second draft for high writers or low writers. High writers who reviewed high quality texts (M = 2.6, SD = 0.2) produced texts of the same quality to high writers who reviewed low quality texts (M = 2.6, SD = 0.2), t(91) < 1, p = .73. Low writers who reviewed high quality texts (M = 2.3, SD = 0.2) produced texts of similar quality to low writers who reviewed low quality texts (M = 2.2, SD = 0.2), t(94) < 1, p = .72. These results indicated that providing feedback helped all participants equally on the revision task. Similarly, all participants benefited equally regardless of the quality of the text reviewed for each of the specific reviewing dimensions (e.g., flow, logic).

Condition effects on future writing. For high writers, there were no significant main effects or interactions on the quality of the second writing assignment. High writers who reviewed high quality texts ($M = 2.6$, $SD = 0.5$) produced texts of similar quality to high writers who reviewed low quality texts ($M = 2.5$, $SD = 0.5$), $t(62) = 1.2$, $p = .24$. For low writers, there was a significant main effect of text quality—that is, low writers who reviewed high quality texts ($M = 2.1$, $SD = 0.4$) produced lower quality texts than low writers who reviewed low quality texts ($M = 2.3$, $SD = 0.4$), $t(65) = -2.0$, $p = .05$, $d = -0.49$.

Conclusion. Based on these findings, providing feedback appeared to be more important for transfer rather than revision. In addition, providing feedback looked to be equally helpful for high writers regardless of the quality of texts being reviewed. Most interesting were the differences for low writers—that is, reviewing low quality text seemed to be more useful for future writing than reviewing high quality text.

3.3.3. Amount, Features, and Focus of Comments

Impact of Comments. In order to determine the impact of the amount, features, and focus of comments on the benefits of providing feedback, I computed the partial correlations predicting the quality of the second paper for each of the measures controlling for the quality of the first draft of the first paper. While most of the partial correlations were not statistically significant (see Table 8), the presence of solutions was positively related to the quality of the second paper, $p_{SolutionsPaper2quality.Paper1draft1quality}(128) = .19$, $p = .03$. These results indicated that students might benefit the most from providing feedback if they include solutions in their comments.

Table 8: Partial correlations of second paper quality

	r	*p*
Amount of Comments		
length of comments provided	.08	.34
number of comments provided	.14	.11
Features of Comments		
number of problems provided	.07	.44
number of solutions provided	.19	.03
number of localized comments provided	.03	.74
Focus of Comments		
number of high prose comments provided	.12	.19
number of low prose comments provided	.14	.122
number of substance comments provided	.10	.24

[a]Controlling for the quality of the first draft of the first paper

Amount of Comments. Participants did not differ in the amount of comments provided to peers across the conditions. High writers generated a similar number of comments for high quality texts (M = 10.2, SD = 3.5) and low quality texts (M = 11.0, SD = 3.5), $t(89) = -1.0$, $p = .31$. In addition, their comments for high quality texts (M = 185, SD = 80) and low quality texts (M = 205, SD = 82) were similar in length, $t(89) = -1.2$, $p = .24$. Likewise, low writers generated a similar number of comments for high quality texts (M = 11.2, SD = 3.6) and low quality texts (M = 11.5, SD = 3.4), $t(93) < 1$, $p = .75$. In addition, their comments for high quality texts (M = 185, SD = 80) and low quality texts (M = 205, SD = 82) were similar in length, $t(93) < 1$, $p = .81$.

Features of Comments. The amount of most features in the feedback provided did not differ significantly across the conditions. The only feature that differed significantly was the amount of problems described. While high writers described fewer problems for high quality texts (M = 2.3, SD = 1.9) than low quality texts (M = 3.4, SD = 1.9), $t(89) = -2.8$, $p = .01$, $d = -0.59$, low writers described a similar number of problems found in high quality texts (M = 2.8, SD = 1.9) as in low quality texts, (M = 2.6, SD = 1.9), $t(93) < 1$, $p = .49$, $d = 0.14$.

By contrast, high writers did not differ in the amount of solutions generated for high quality texts (M = 2.5, SD = 2.0) or low quality texts (M = 2.9, SD = 2.0), $t(89) < 1$, $p = .39$, $d = -0.18$, and they also did not differ in the amount of localization included for high quality texts (M = 1.2, SD = 1.6) or low quality texts (M = 1.7, SD = 1.6), $t(89) = -1.5$, $p = .13$, $d = -0.32$. Similarly, low writers did not differ in the amount of solutions generated for high quality texts (M = 2.2, SD = 2.1) or low quality texts (M = 2.5, SD = 2.0), $t(93) < 1$, $p = .46$, $d = -0.15$, and they also did not differ in the amount of localization included for high quality texts (M = 1.5, SD = 1.6) or low quality texts (M = 1.4, SD = 1.5), $t(93) < 1$, $p = .84$, $d = 0.04$.

Focus of Comments. Similar to the amounts of features in the provided feedback, the participants differed very little in the amount of focus of the feedback provided. Only the amount of high prose comments differed significantly across the conditions. While high writers described fewer high prose issues for high quality texts (M = 2.3, SD = 1.4) than low quality texts (M = 3.0, SD = 1.4), $t(89) = -2.4$, $p = .02$, $d = -0.51$, low writers described the same number of high prose issues for both high quality texts (M = 2.2, SD = 1.4) and low quality texts, (M = 2.2, SD = 1.3), $t(93) < 1$, $p = .89$, $d = 0.03$.

By contrast, high writers did not differ in the amount of low prose comments generated for high quality texts (M = 0.6, SD = 1.0) or low quality texts (M = 0.8, SD = 1.0), $t(89) = -1.0$, $p = .33$, $d = -0.20$, and they also did not differ in the amount of substance comments generated for high quality texts (M = 1.7, SD = 1.4) or low quality texts (M = 2.0, SD = 1.4), $t(89) = -1.1$, $p = .27$, $d = -0.23$. Similarly, low writers did not differ in the amount of low prose comments generated for high quality texts (M = 0.7, SD = 1.0) or low quality texts (M = 0.8, SD = 1.0), $t(93) < 1$, $p = .73$, $d = -0.07$, and they also did not differ in the amount of substance comments generated for high quality texts (M = 1.6, SD = 1.5) or low quality texts (M = 1.6, SD = 1.4), $t(93) < 1$, $p = .88$, $d = 0.03$.

Conclusion. Participants differed very little in the amount, features, and focus of provided comments across the conditions. High writers described more problems and focused more often on high prose issues when reviewing low quality papers. However, these differences did not seem to affect the quality of revisions or future drafts; the quality of high writers' second draft and second paper were similar after reviewing high quality texts or low quality texts.

By contrast, low writers did not differ in amount, features, or focus of comments generated for high quality texts or low quality texts. Similar to the high writers, this difference

did not extend to the quality of future writing—that is, the quality of low writers's drafts were not similar after reviewing high quality text and low quality text. In fact, low writers produced poorer quality documents if they had reviewed high quality texts.

3.4. GENERAL DISCUSSION

3.4.1. Summary of Results

The purpose of the current study was to understand to what extent the quality of the text being reviewed and the student's ability affected how much students could learn from reviewing peers' texts. Analyses focused on the draft quality of revised texts and future writing, as well as the amount, features (i.e., problem, solution, and localization), and focus (i.e., high prose and substance) of the comments provided. High prose comments focused on high-level writing issues, such as "To me, rather than posing an argument, it just seems as though you are retelling the study. Take a side and back it up using details from the articles and also bring in outside information that will help support your position, just make sure it is revelant the the [sic] articles." Substance comments focused on the inclusion of accurate content in the paper, such as "Explain in more depth why causal conclusions cannot be made from the correlations." Overall, providing feedback seemed to be more important for future writing tasks rather than the revision task. This benefit differed for high writers and low writers.

For high writers, providing feedback was equally beneficial regardless of the quality of the reviewed text. Furthermore, while the amount and most features/focus of comments did not differ depending on the quality of the text reviewed for high writers, they described more

problems and focused more often on high prose issues when reviewing low quality papers. Despite these differences, there was not an observable benefit of reviewing low quality papers over high quality papers on future writing.

For low writers, providing feedback was useful only if reviewing low quality texts. In addition, low writers' provided comments did not differ in amount, features, or focus, which indicated that low writers could benefit equally from reviewing high quality texts and low quality texts. However, this finding contradicts the draft quality outcome.

3.4.2. Classic Model of Cognitive Processes in Revising

Why were these overall findings observed? A theoretical understanding of the benefits of reviewing peers' texts can be built around the components of the revision task with which students struggle most. Flower and Hayes (1981) identified reviewing as one of the three cognitive processes of writing. At this time, however, very little was known about the process of reviewing and how the two sub-processes of reviewing (i.e., evaluation and revision) worked. In a later extension of the model, Flower, Hayes and colleagues further described the cognitive processes involved in revision—these processes included task definition, evaluation, problem detection, problem diagnosis, and strategy selection (Flower, et al., 1986; Hayes, et al., 1987). They further described expert-novice differences, which could be useful in understanding where in the revision task students have the most difficulty.

As first component of the model, a writer must *define* the task—that is, the writer begins with an expectation about the goals, constraints, and criteria for the task. These expectations may be retrieved from the writer's knowledge about the genre or expectations may develop from specific instructions about the task. The task definition takes the form of an internal

representation of the task, which is critical for detecting problems. Two aspects of the task defining process differ between experts and novices: 1) novices are less likely to define the task beyond explicit instructions, and 2) novices have a shallow understanding of the task, which results in a task definition that involves fewer goals, constraints, and criteria—especially ones that focus on global issues—that are used during evaluation to detect a fewer number of problems.

Evaluation takes place as a writer reads the text-produced-so-far. While reading, the writer has certain goals in mind. At the simplest level, the writer may read to satisfy the goal of comprehension. While reading, the writer both creates an internal representation of the text and compares that representation to the task definition to determine whether the text meets the set expectations. As the goal changes, the task definition is also revised to incorporate additional constraints and criteria. Therefore, a writer's ability to evaluate the text is directly affected by the writer's ability to define the task.

Problem detection occurs when the writer experiences a negative evaluation—that is, when the representation of the text does not meet the expectations set by the writer. Even with the simplest goal in mind (i.e., reading for comprehension), the writer may detect problems as holes in their representation of the text become apparent. By increasing the level of the goal from reading for comprehension to reading for evaluation or defining problems, the writer is able to detect an increasing number of problems—that is, when the representation of the text does not match the writer's expectations (i.e., task definition), the writer becomes aware of a problem with the text. Problem detection could be disrupted in two ways. First, the writer may develop a poor representation of the text—that is, frequently gaps in the text are automatically filled in resulting in an inaccurate representation of the text (i.e., a representation that has too few errors).

This problem commonly occurs among both experts and novices. Many experts are aware of this problem, so they try to use strategies that eliminate the formation of the text's meaning that could lead to automatic gap filling. Unfortunately, these strategies only help with detection of low prose issues, as it is difficult to detect global problems without a complete representation of the text. Second, the writer may have a poor representation of the task—that is, by using an impoverished task definition, only a narrow range of problems could be detected. This problem affects novices more than experts because the novices use a less detailed task definition.

Problem detection is only the first step in successful revision. Once a problem is detected, the writer needs to know how to fix the problem. Flower and colleagues defined this aspect of revision as *diagnosis*—that is, the ability to describe the problem, why it is problematic, and how to fix it. Diagnoses range on a continuum from 'simple detects' or 'ill-defined problems' (i.e., the writer is only aware that something is not right—there is a lack of information about the nature and locus of the problem) to 'fully explicit' or 'well-defined problems' (i.e., the writer recognizes the problem and knows specific rules about how to fix this particular problem—usually only occurs with low prose issues where explicit rules can be applied, such as grammar and spelling). When a writer is capable of defining the nature and locus of the problem, the writer is more likely to retrieve additional information about how to fix the problem detected. Novices are limited in their ability to diagnose problems because they have a small repertoire of problem representations. They both know of fewer problems and have only a shallow understanding about those problems. Therefore, when novices detect a problem, they are less likely to recognize it as something they know, and they are not likely to retrieve more information about how to fix the particular problem.

Finally, the writer must *select a strategy* for revision. Both problem detection and diagnosis are critical for a writer to choose to revise. If only detection occurs, the writer is not likely to understand the nature of the problem or how to fix it. In these instances, the only option the writer has to fix the problem is to rewrite the text in hopes that the new text would no longer contain the originally detected problem. Because rewriting occurs when the writer does not really understand the problem trying to be fixed, the initial problem may continue to exist after the rewrite and/or new problems may occur. By contrast, if a writer is able to both detect and diagnose the problem, the additional information could inform the writer how to revise the problematic text. Since revising occurs when the writer understands where the problem occurred and how to fix it, the problem is more likely to be resolved successfully.

Students are especially likely to struggle with three components of the revision task. First, their task definitions are likely to lack many important features that would limit their ability to evaluate the text. Second, they are also less likely to detect problems with the text. Finally, they are less likely to retrieve additional information about the problem, including how to fix the problem.

3.4.3. Learning Opportunities during Peer Review

With a theoretical model in hand for the cognitive processes involved in writing, I can now discuss the learning opportunities found in peer review in terms of different possible learning mechanisms that could apply to those processes. Further, the learning mechanisms could provide a framework for understanding the ways in which writing ability, text quality, and feedback source might moderate those learning opportunities.

Typically, students complete two tasks while reviewing peers' papers—first, students rate the quality of the paper based on the specified grading criteria, and then they offer the peer feedback about the major problems and how to fix them. For each of these tasks, students complete a number of subtasks—each a possible learning opportunity for the student to improve their ability to revise while writing. Learning during each of the subtasks is likely to occur by different mechanisms, and thus be influenced by different situational factors (see Figure 3).

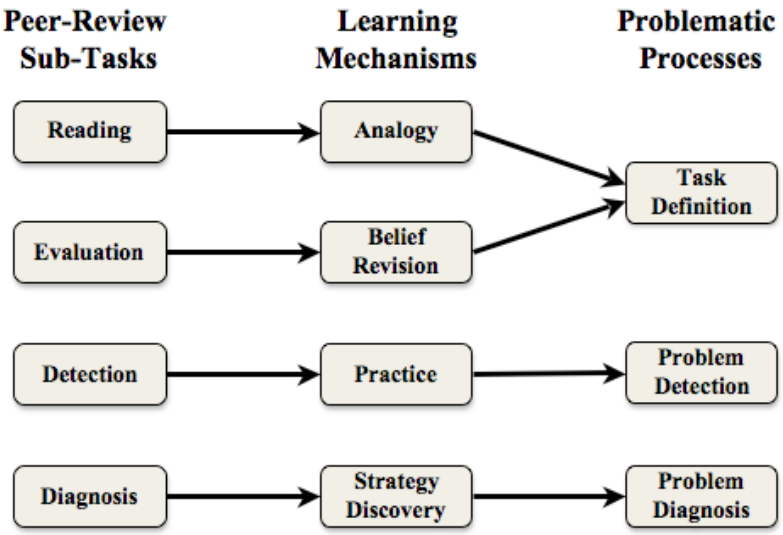

Figure 3. Learning opportunities and possible learning mechanisms during peer review

The first peer review subtask a student must complete is reading. As mentioned previously, several goals could be applied to the reading task: reading for comprehension, reading for evaluation, and reading to define problems (Flower, et al., 1986; Hayes, et al., 1987). While reading other peers' texts, the student creates a mental representation of the text. In addition to representing the argument of the paper (i.e., take-home message), this representation

may also include information about the task—that is, the peers' papers could serve as models of success and/or failure.

As novice writers, students' task definition is likely to be shallow. Observing these models could help the students further develop their task definition through revised criteria that include additional instances of problems. In addition, by reviewing multiple peers' papers, students may formulate a more generalizable task definition (i.e., new broader criteria or goals). In comparing multiple models, students could extract details about the task that is common across the models. This representation from multiple models is likely to be context independent, which could facilitate retrieval in a variety of future writing tasks that differ from the current writing task. Having a richer task definition could help the students on future writing endeavors by enabling them to detect and resolve a greater quantity of problems with their text while they are writing.

This learning of task definitions by reading can be thought of an instance of learning by analogy (Catrambone & Holyoak, 1989; Gentner, Loewenstein, & Thompson, 2003; Holyoak, 1985). While reading a peer's paper, the student could draw analogies between the text representation and his or her own current task definition. If the two representations do not match, the student must resolve the dissonance. There are at least two possible options: first, the student might decide that either there is a problem with the peer's text that should be described in feedback, or second, the student might believe that the problem lies within his or her own task definition, which could result in an update to the task definition. Similarly, before revising one's own paper, the student could draw analogies the text representation and his or her own current task definition. Again, if these representations do not match, the student might decide there is a problem with the text and revise accordingly, or the student might decide that the current task

definition was not appropriate for the task and update the task definition. Finally, during future writing assignments, the student could draw analogies between encoded model representations of writing (e.g., peers' texts that were read for a prior peer review assignment) and their own text representation. If these representations do to match, the student would need to decide whether the model was not appropriate resulting in a task definition update or whether a problem exists in one's own text resulting in revision of the text. In any of these cases, these analogical processes could impact the student's own task definitions. However, the question of whether the reader actually will draw these analogies remains. The learning by analogy literature suggests that learners often struggle to retrieve relevant prior examples when they need them because superficial elements are most salient in their memories (Forbus, Gentner, & Law, 1995), and this effect is stronger for more novice students (Novick, 1988).

The next peer review subtask that the student must complete is evaluation. This subtask can take on several forms. The most obvious is the evaluation of the overall quality of the paper, which is necessary for the student to determine which quality rating is most appropriate. While many instructors provide grading rubrics to their students in order to make their expectations explicit, students might see these guidelines as an arbitrary opinion of the instructor as opposed to guidelines about good writing. Even when the student acknowledges that the grading rubric is directly related to qualities of good writing, the student may have only a shallow understanding about what that actually means. By acting as the evaluator, the student could learn through experience how the grading rubric actually measures varying degrees of quality. As a result, the student's task definition could become richer, again allowing them to write higher quality texts in the future.

Another form of evaluation that students should complete is evaluating the severity of the problems detected. While reading a peer's text, many problems are likely to be detected. The student must decide which ones to include in the feedback. Ideally (if such instructions are provided and they are followed), the student would choose to describe the major problems of the paper. Therefore, when a problem is detected, it should be compared with the other problems detected to determine if which one is more important. In doing so, the student gains additional information about which types of problems affect the paper the most. This information could be useful when prioritizing one's time while revising.

This effect of evaluation changing one's task definition could be thought of as an instance of belief revision. While there are several different theories about belief revision and conceptual change (for a review, see Ohlsson, 2009), most researchers would agree that the more discrepant instances appear to influence beliefs more strongly that less discrepant instances (Chan, Burtis, & Bereiter, 1997). Two anecdotal examples illustrate this effect. For most academic writing assignments, students are likely to just focus on demonstrating their knowledge of the content rather than clearly communicating a coherent argument. As one professor observed, students in his physics class frequently write formal lab reports like legal disclaimers—that is, they believe that if every detail about the lab was included in the report, they would receive full credit. Unfortunately, these papers became very laborious to grade. As noted by this professor, one of the major benefits of requiring peer review was that students realized reading papers that included too much information was painful and really unnecessary. A similar reaction was found in a graduate-level course about reasoning and problem solving. After reviewing peers' papers, students were asked to reflect on what they learned about writing. One student stated, "After reviewing my first set of papers I had a better idea of the level of background detail that should

be included in a paper to make it understandable to someone who is not very familiar with the topic. Also, I discovered the value of well chosen sub-headings and a clearly stated thesis. Finally, I concluded that style really does matter. Some papers were very dry and it seemed as though the authors were uninterested in the topic. I think it is important to find a way to get interested at least while writing, or else the idea and the overall paper suffers, no matter how insightful it may be." As these examples demonstrated, poor quality writing often constituted the discrepant instance and led to a revision to the task definition. Thus, if students benefited from peer review via evaluating paper quality, then I would expect that reviewing poor quality texts would generally lead to more improvements in student writing than reviewing high quality texts.

The third peer review subtask that the student must complete is problem detection. This subtask could be useful when determining which quality rating is most appropriate, and it is especially important for writing feedback. Without instructional support, students are likely to detect relatively few problems because their task definition is poorly developed. In particular, without additional prompts, students are likely to focus on low prose issues, such as grammatical and spelling errors. However, after only brief training on global problems, students are capable of detecting more problems (Wallace & Hayes, 1991). Therefore, in peer review, providing detailed instructions that describe typical problems to look for could help the students detect the relevant problems for their feedback. Through practice detecting new types of problems, students could strengthen their ability to detect these problems, resulting in faster and more efficient retrieval of information about these problems while writing in the future.

This kind of learning effect is a simple skill practice effect, for which there are many models (Anderson et al., 2004; Logan, 1988; Newell, 1994; Newell & Rosenbloom, 1981). In general, the more problems that are detected, the more benefits come from this kind of learning

mechanism (Anderson, et al., 2004). Under this theoretical framing, students would likely benefit more from practice if the texts being reviewed had many problems. Therefore, if students benefited from peer review via general problem detection, then I would expect that reviewing poor quality texts (assuming poor quality texts present more problems to detect) would lead to more improvements in student writing than reviewing high quality texts. On the other hand, practice that is focused on particular areas of weakness has been shown to be more efficient in producing learning gains (Anderson, Corbett, Koedinger, & Pelletier, 1995). Under this framing, students might learn more from reviewing texts that suffer from similar problems to their own writing problems. Thus, if students benefited from peer review via focused problem detection, then I would expect an interaction between writer ability and text quality—that is, reviewing poor quality texts would lead to more improvements in student writing than reviewing high quality texts for low-ability writers, and reviewing high quality texts would lead to more improvements in student writing than reviewing low quality texts for high-ability writers.

The final peer review subtask that the student must complete is problem diagnosis. Once the student has chosen which problems to include in the feedback, information about those problems (e.g., what is the problem, why is it problematic, how to fix it) must be described. However, this subtask could be the most difficult for students because novice writers tend to have shallow understanding of writing problems—that is, they may be able to recognize the problem, but they may not be able to explain why it is problematic or how to fix it. At this point, the student has two options: 1) ignore the problem and decide not to include it in the feedback or 2) invent new solutions that were not previously in the student's prior set of revision techniques.

Through the second option, a strategy discovery process, students can further develop their problem diagnosis repertoire by increasing the number of problems that they are able to

understand as well as deepening their understanding of those problems. In future writing situations, the students should be able to fix more problems resulting in higher quality writing. Researchers observed that situations in which old strategies were particularly inefficient or ineffective lead to the discovery of new strategies (Siegler & Stern, 1998). These results suggested that students could discover solutions to problems in situations where they do not already possess a solution. Therefore, if students benefited from peer review via diagnosing the problems, then I would expect that reviewing texts of a quality that were of equal or higher quality would lead to more improvements in student writing than reviewing texts of lower quality than their own.

How did these various learning mechanism based expectations for ability-related outcome differences compare to the results that were found? While the inferences are tentative due to the indirect reasoning required to connect data patterns to theory, the following connections are apparent:

1. Because low writers wrote higher quality second papers after constructing feedback for low quality texts, learning might have occurred via two possible mechanisms: evaluating the draft quality of the paper (i.e., belief revision) and/or detecting writing problems (i.e., general practice or focused practice).
2. Because high writers described more problems and focused more often on high prose issues when reviewing low quality papers, learning might have occurred via detecting problems in general (i.e., general practice). However, these differences did not seem to extend to the quality of the drafts—that is, the quality of high writers' second draft and second paper did not differ after reviewing high quality texts or low quality texts. Several possible reasons could explain this lack of an effect on draft quality: first,

high writers might only improve their writing ability by describing solutions (i.e., the only feature that significantly predicted future draft quality), and second, high writers might be practicing detection of problems in which they already deeply understand.

3. Because low writers did not differ in amount, features, or focus of comments generated for high quality texts or low quality texts, learning might have occurred via diagnosing new problems (i.e., strategy discovery). There were likely to be problems in both high quality texts and low quality texts that a low writer did not know how to solve. By forcing the student to articulate a solution, a new strategy must have been discovered. However, this difference did not extend to draft quality—that is, low writers did not benefit equally from reviewing high quality texts and low quality texts. In fact, low writers produced poorer quality documents if they had reviewed high quality texts. Perhaps, the problems in the high quality text were too far from the low writers ability, and the solutions they discovered could have been ineffective and potentially harmful.

3.4.4. Caveats & Future Directions

The current study did not directly test the possible learning mechanisms. Therefore, future research should systematically test each of the possible mechanisms. By manipulating the possible mechanisms (e.g., whether students are instructed to compare/contrast model texts), one could better understand how students might learn from the reviewing task, and thus possibly maximize the benefits of reviewing peers' texts.

In addition, future research should develop a tool that could automatically identify which specific writing problems a student struggles with. This tool could be used to assign students to

peer review groups that could maximize their practice in detecting and diagnosing problems that they do not fully understand.

3.4.5. Implications for Classroom Practice

Overall, both high writers and low writers seem to benefit equally from reviewing high quality texts and low quality texts. There were very few differences in the amount, features, and focus of the provided feedback, which did not affect the students' ability to revise. Thus, randomly assigning students to peer review each others' papers does not seem like a problematic strategy. These results also lead to the tentative implication that students may benefit more if working with similar ability peers. Therefore, if an instructor feels particularly cautious about using peer review, assigning students to review peers of the same ability level could avoid a possibly less useful case (i.e., low writers reviewing high quality texts). In addition, advances in intelligent systems research could also be used to help students focus the reviewing on the relevant problems with which they also continue to struggle. The current research provides some information on which kinds of feedback appear to be most helpful for different student types.

4.0 GENERAL CONCLUSION

4.1.1. Theoretical Contributions

I examined a range of mechanisms that might explain the impact of reviewing peers' texts and receiving peer feedback on writing: analogy, belief revision, general and focused practice, strategy discovery, amount of comments received by features and types of focus, and implementability of comments by features and types of focus. Different mechanisms were expected to impact improvements in the current document (i.e., performance) and quality of future documents (i.e., transfer)—each of which are further discussed below.

Performance. In the first study, I explored how receiving peer feedback affected the students' revision process. Only three of the coded features and types of focus of peer feedback had a significant effect on the overall revision process: localized comments negatively impacted the quality of the revised document, and high prose and substance comments positively impacted the quality of the revised document. The negative effect of localized comments was surprising and contrary to typical advice given to students in practice and some prior research (e.g., Nelson & Schunn, 2009). Perhaps, being provided the location of the feedback makes revising the document easier for the students—they would be able to go to the specified location and modify the text. However, such localized revisions might disconnect that text from the surrounding prose, which could result in a less coherent text. Unraveling this piece will likely lead to improved theoretical models for how peer feedback affects revisions.

In addition to some new findings regarding the general impact of feedback features and types of focus, I also found that the extent to which this highlighted feature and types of focus

affected the revision process differed for high writers and low writers. High writers were equally likely to implement localized comments and high prose comments from high-ability sources and low-ability sources; furthermore, they were also equally likely to receive localized comments and high prose comments from high-ability sources and low-ability sources. More importantly, high writers were more likely to implement substance comments if provided by high-ability sources, and they were also more likely to receive substance comments from high-ability sources than low-ability sources. Despite these differences, the texts revised based on feedback from high-ability sources were not higher in quality than the texts revised based on feedback from low-ability sources.

For low writers, the feedback source appeared to have more influence on the important feature and types of focus. Low writers were more likely to implement localized comments if provided by low-ability sources; however, these revisions did not greatly affect the overall draft quality because they were more likely to receive localized comments from high-ability sources rather than low-ability sources. Similarly, while low writers were more likely to implement substance comments if provided by low-ability sources, they did not benefit from these revisions because they were more likely to receive substance comments from high-ability sources than low-ability sources. Finally, low writers were more likely to implement high prose comments if provided by low-ability sources. By receiving the same amount of high prose comments from both high-ability sources as low-ability sources, it was possible that low writers benefited slightly more from receiving feedback from low-ability sources. However, these differences did not affect the overall draft quality—low writers who received feedback from high-ability sources produced revised drafts of similar quality as those who received feedback from low-ability sources.

Overall these complex interactions of writing ability and feedback source highlight the importance of separately considering the multi-layered nature of effects of peer feedback on revision: amounts of feedback amounts, implementability of feedback, and revision quality. To maximize students' ability to revise using peer feedback, improvements need to be made to all three layers—that is, increases in the amount and implementability of specific feedback that is most likely to lead to draft quality improvements. Further, the nature of each layer may change according to a variety of situational factors, including the skill set of the feedback provider and the feedback receiver.

Transfer. In the second study, I examined how reviewing peers' texts benefited students. Overall, this reviewing task seemed to help students with future writing rather than their revisions—however, it was unclear whether the reviewing task did not help at all or helped all students equally. Further, based on a careful task analysis of revision and learning opportunities found in peer review, the more plausible mechanisms for learning from reviewing seemed to differ for high writers and low writers. During the reviewing task, high writers practiced describing problems (especially those focused on high prose issues) more often when reviewing low quality papers than high quality papers. Nevertheless, this additional practice did not seem to be helpful—high writers who reviewed low quality texts produced papers of the same quality as those who reviewed high quality texts. These results suggested that for high writers general practice might not be enough; they might have needed either practice that was more focused on particular problems or practice that involved describing solutions.

During the reviewing task, low writers' comments did not differ in amount, features, or focus, which suggested that these students could benefit equally from reviewing high quality texts and low quality texts. In both cases, there were likely to be problems with which the low

writers continued to struggle; thus the reviewing task provided them with an equal number of opportunities to discover solutions to the new problems found in both high quality texts and low quality texts. Surprisingly, analyses of future draft quality did not further support this conclusion. The quality of the future paper differed depending on the quality of the texts reviewed—that is, low writers who reviewed low quality texts wrote better future papers than those who reviewed high quality texts. These results indicated that while low writers might have equal opportunities to discover solutions to new problems detected in both high quality texts and low quality texts, the solutions discovered while reviewing high quality texts might have been ineffective or harmful.

4.1.2. Pragmatic Contributions

Overall, the quality of the text to be revised and the feedback source seemed to have relatively little influence over students' revision process. However, the few significant differences found in the two studies provided evidence towards a tentative claim: the benefit of peer review could be maximized with individualized instruction—that is, high writers might benefit the most from peer review if grouped with other high peers, and low writers might benefit the most from peer review if grouped with other low peers.

Students seemed to benefit the most from reviewing peers' texts when given an opportunity for focused practice. To maximize the benefits of peer review, students could be assigned to review texts written by peers of the same ability and encouraged to describe both problems and solutions about high prose and substance issues. In return, they might receive more comments about high prose and substance issues, which if encouraged to use during revision could lead to greater improvements in draft quality. Moreover, by articulating solutions for

problems with which they continue to struggle, they might discover new solutions. Using these new solutions in future writing could lead to higher quality papers.

The quality of the text to be revised and the feedback source did have some influence over students' revision process. However, based on correlations of the relative distribution of comment types, high-ability students and low-ability students received very similar types of comments, $r(4) > .99$, $p < .001$ and implemented similar types of comments, $r(4) = .91$, $p = .01$ (see Figure 4). This high degree of overall similarity in commenting content suggests that students of varying ability levels would likely benefit relatively similarly from peer review. This observation has been made before (Patchan, et al., 2009), but nonetheless may be surprising to many instructors and students. In particular, they are often particularly concerned about how helpful low writers could be as reviewers and whether providing feedback to low peers could help high writers at all.

Probably the most worrisome case is when low-ability students are assigned to review each other's papers. Are low-ability students capable of helping each other? First, low ability peers provided very similar feedback in comparison to feedback provided by high-ability peers, which led to revised papers of comparable quality. In addition, I observed that low-ability students are likely to benefit more from reviewing low-ability peers' papers than reviewing high-ability peers' papers. Thus, low-ability students appear to be capable of helping each other.

Instructors also worry about low-ability students being assigned to review high-ability peers' papers. For this case, there are two main concerns: 1) are low-ability students capable of helping high-ability peers, and 2) do low-ability students benefit when trying to help high-ability peers? First, high-ability students received comments from low-ability peers that were very similar in character to the comments from high-ability peers. For high-ability students, there

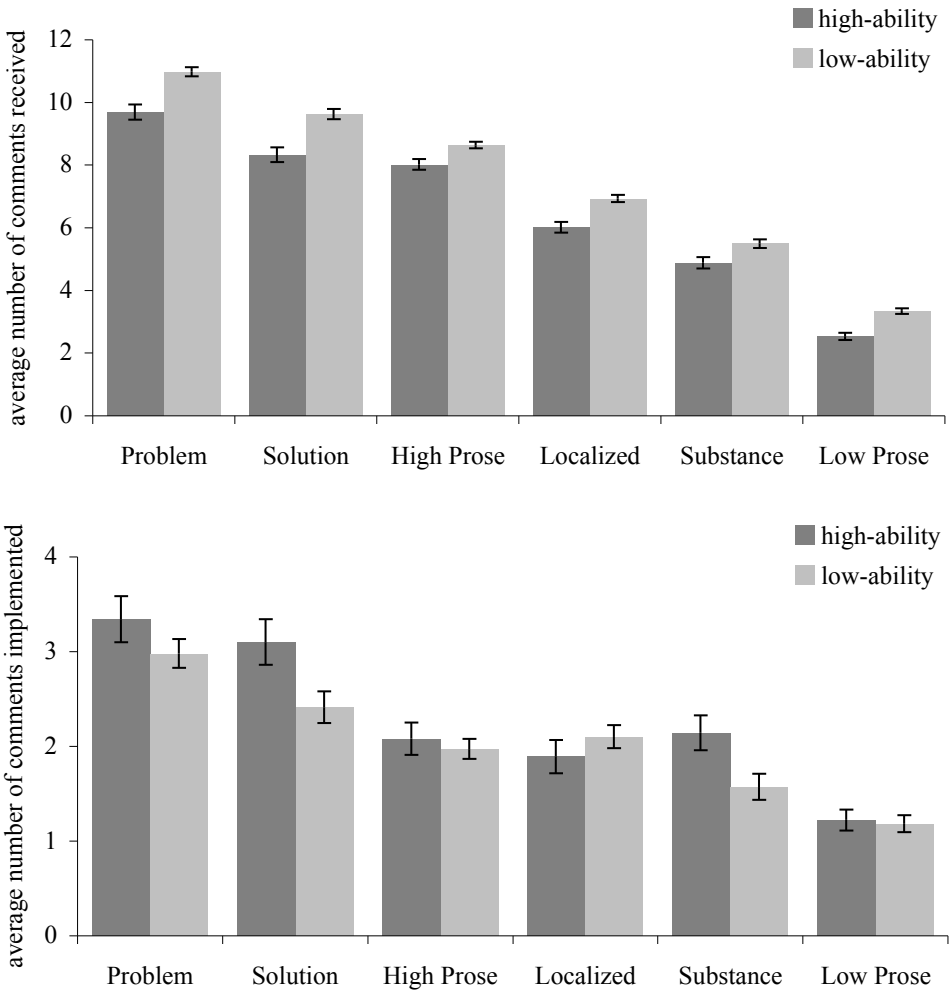

Figure 4. Relative distributions of comment types

were no differences in quality of revised drafts, regardless of who provided the feedback. Thus, low-ability students appear to be capable of helping high-ability peers. One possible drawback in this case could be that low-ability students do not seem to benefit as much from reviewing high-ability peers' texts as they do from reviewing low-ability peers' texts. The low-ability students who reviewed high-ability peers' texts produced future papers of lower quality than those who

reviewed low-ability peers' texts. Therefore, low-ability students do not appear to benefit when trying to help high-ability peers. If possible, instructors should try to avoid this case when assigning peer reviews.

Finally, there may be some concern in assigning high-ability students as reviewers of low-ability peers' texts. There is no doubt that high-ability students could help their low-ability peers, but some may wonder about how helpful the assignment is for the high-ability students. Do high-ability students benefit from reviewing low-ability peers' papers? I observed that high-ability students described more problems—especially ones that focused on high prose issues when reviewing low-ability peers' texts. While this extra practice did not see to lead to additional learning gains, the high-ability students who reviewed low-ability peers' texts produced new papers that were of comparable quality to those who reviewed high-ability peers' texts. Therefore, high-ability students appear to benefit equally from reviewing low-ability peers' papers.

4.1.3. Methodological Contributions

In the current studies, a number of new methodological approaches to studying peer review were used. First, our experimental design allowed me to tease apart effects of providing feedback from effects of receiving feedback, in contrast to other designs used in prior research that focused either on only receiving feedback or providing feedback (e.g., Cho & Schunn, 2007; Wooley, et al., 2008). Also, unlike the prior research designs, the variables in the current study were manipulated completely behind the scenes without students' awareness of the manipulation process. Furthermore, this design also allowed me to observe whether there were any interactions between writer ability, the quality of the text to be reviewed, and feedback source. Observing

these possible interactions is especially important because they could help explain how students learn from peer review and provide more information about how to best develop peer review assignments.

Second, a reliable and valid coding rubric was developed to measure the quality of the revisions. In prior research, either a more qualitative approach was used to examine revision in small datasets (e.g., Faigley & Witte, 1981) or researchers focused on the probability of feedback implementation while ignoring revision quality (e.g., Nelson & Schunn, 2009). This new rubric helped extend the prior research by determining which comments would actually lead to increases in draft quality.

Third, I utilized a new analysis technique that allowed me to better understand how writer ability and feedback source affected the various stages in the revision process. With this new technique, I was able to compare which types of comments were most important (i.e., influenced draft quality the most) to which types of comments actually differed in their implementability and amount received. Overall, this analytic technique provides a systematic analysis of each layer in isolation in addition to providing a method for integrating the layers together. In combination, the multi-layer information derived from this approach can be used to explain why there were no differences in final draft quality while there are differences in components of the feedback. Thus, I can suggest practical areas of focus for further improvements.

Finally, I extended prior research about peer review by examining not only performance on multiple drafts but also performance on future writing tasks. By contrast, prior research has either examined revision activity as a direct performance effect (e.g., Tseng & Tsai, 2007) or observed the cumulative change with extended peer review without being able to examine components of peer review on transfer (e.g., Lundstrom & Baker, 2009). By including an

additional writing assignment on a new topic, I was able to determine whether the effects observed on revision transferred to future writing and whether missing effects on revision showed up later on future writing.

4.1.4. Caveats & Future Directions

First, the assignment context of this study could have influenced these results. Specifically, the focus of the feedback may have been influenced by the detailed rubric that focused on high-level writing issues. The three dimensions chosen provided scaffolding to students so that they would focus on both high prose issues (such as having clear main ideas and smooth transitions between ideas) and content issues (such as providing appropriate support and considering possible counter-arguments). Without this kind of support, novice writers tend to focus more on lower-level issues (Wallace & Hayes, 1991). If a detailed rubric was not offered to the students, two possible results may occur. High-ability sources may focus on high-level issues while the low-ability sources may focus on low-level issues, or there may be no differences between high-ability and low-ability sources—all students may focus equally on low-level issues. Therefore, receiving a detailed rubric might increase the possible benefits of reviewing peers' texts.

Furthermore, this assignment required students to generate end comments (i.e., comments written outside of the paper) rather than marginalia (i.e., comments written in the margins of the paper). By generating end comments, students are more likely to focus on global issues, whereas students who provide only marginalia may be more likely to focus on low prose issues. Also, given the unlimited space available for end comments, in contrast to the limited space available in the margins, students may be more inclined to describe problems and possible solutions in

more depth. Thus, generating end comments might have a greater impact on students' ability to detect and diagnose problems.

Similarly, the draft quality measure could have also influenced the results. The current measure was chosen because it closely matched the goal of the course and the writing assignment. Theoretically, other standardized measures, such as SAT/GRE writing, might have revealed different results because there is more emphasis on lower-level problems in these measures. Under this possibility, because students commonly focus on lower-level issues during revision, larger gains between first and second drafts might be observed. Practically speaking however, most of the feedback received did not focus on lower-level issues, so these gains would not likely differ between conditions.

Writer ability was measured using an indirect measure (i.e., composite of self-reported SAT verbal, SAT writing, and grades in freshman composition courses). For research purposes, future studies could benefit from using more direct measures of writer ability. For pragmatic purposes, these indirect measures have great value (i.e., easy to collect at the beginning of the semester). Therefore, future research should validate which indirect measures are most appropriate or develop ways to automatically determine writer ability from direct measures of writing without the need for human evaluation.

In the current study, implementation was coarsely defined—if a writer attempts to make a change in the paper in regards to something mentioned in the comment, it was considered implemented. A coarse implementation code was used because coders experienced difficulty in attempts to code finer-grain types of implementation (i.e., attaining reasonable reliability rates were not possible). The main cause for difficulty in fine-grained implementation coding was that it was often unclear why a student chose to make a certain revision (i.e., was it in response to a

particular piece of feedback or was it for some other reason), and similarly, it was unclear why students seemed to ignore the majority of the feedback. Future research should examine more closely how students choose which comments to implement (i.e., use of think-aloud or post-hoc reflection). In addition, future research could observe whether students are more likely to implement one reviewer's comments over another reviewer (i.e., can students distinguish a good reviewer from a poor reviewer; does the order of reviewers matter—for example, perhaps the first reviewer's comments are more likely to be implemented regardless of quality?).

In addition, future research should consider using an extreme group comparison to see whether the current results extrapolate to even higher writing ability or to even lower writing ability than was studied here. Possibly, the current results are muted by the presence of many middling ability students.

Furthermore, other populations should be examined. The current population consisted of students with a fairly wide range of ability levels. However, this wide range is not found at all institutions. For example, top tier universities tend to have populations in which the range of ability would be narrow and skewed towards the higher end of the range. This type of population could produce different results. For example, in these cases, the low writers might not be that different from the high writers; therefore, few significant differences would likely be found. Yet future research with these higher writing ability populations is still important because it could help explain more fine-grained developmental differences across the writing expertise continuum.

Finally, the current study focused on writing ability and learning to write (i.e., improvements between drafts and on future writing). However, many researchers are interested in content knowledge's role in writing and writing to learn—that is, using writing assignments to

improve content knowledge. In addition to this question of what should be learned (writing ability or content knowledge), prior content knowledge differences among students can also be an important factor in the writing process. If the writer is familiar with the content, then more cognitive resources can be allocated to the writing process. If the writer is less familiar with the content, then there are fewer cognitive resources to devote to the writing process. This sharing of cognitive resources could negatively impact the writer's ability to write well. All of the students in the current study were less familiar with the content—that is, they were working on their writing assignment within a week of being introduced to the content.

Many researchers have investigated the benefits of writing on content knowledge (Bangert-Drowns, Hurley, & Wilkinson, 2004). In the current study, high writers received and implemented more substance comments from other high-ability sources. This result suggests that receiving feedback from high-ability sources may increase high writers' content knowledge more than if feedback was received from low-ability sources. In contrast, low writers were more likely to receive substance comments from high-ability sources, but they were less likely to implement these comments. Therefore, it was unclear whether they would gain more content knowledge from receiving feedback from high-ability peers or low-ability peers. Future research should focus on these interactions—how different levels of content knowledge affect the quality of writing and how writing about less familiar content improves knowledge of that content.

5.0 REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review, 111*(4), 1036-1060.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences, 4*(2), 167-207.

Anrum, R., & Roksa, J. (2010). *Academically adrift: Limited learning on College Campuses*. Chicago: University of Chicago Press.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29-58.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming Contextual Limitations on Problem-Solving Transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147-1156.

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15*(1), 1-40.

Charney, D. H., & Carlson, R. A. (1995). Learning to Write in a Genre - What Student Writers Take from Model Texts. *Research in the Teaching of English, 29*(1), 88-125.

Cho, K., & Schunn, C. (2010). Developing Writing Skills Through Students Giving Instructional Explanations. In M. K. Stein & L. Kucan (Eds.), *Instructional Explanations in the Disciplines* (pp. 207-221): Springer.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409-426.

Cobb, K. L. (1999). Interactive videodisc instruction with undergraduate nursing students using cooperative learning strategies. [Article]. *Computers in Nursing, 17*(2), 89-96.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal, 19*(2), 237-248.

Crespo Garcia, R., & Pardo, A. (2010). *A supporting system for adaptive peer review based on learners' profiles*. Paper presented at the Workshop on Computer Supported Peer Review in Education.

Crowley, K., & Siegler, R. S. (1993). Flexible strategy use in young children's tic-tac-toe. *Cognitive Science, 17*(4), 531-561.

Day, E. A., Arthur, W., Bell, S. T., Edwards, B. D., Bennett, W., Mendoza, J. L., et al. (2005). Ability-based pairing strategies in the team-based training of a complex skill: Does the intelligence of your training partner matter? [Article]. *Intelligence, 33*(1), 39-65.

De Lisi, R., & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 3-37). Mahwah, NJ: Lawrence Erlbaum Associates.

Faigley, L., & Witte, S. (1981). Analyzing Revision. *College Composition and Communication, 32*(4), 400-414.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication, 32*(4), 365-387.

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. *College Composition and Communication, 37*(1), 16-55.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*(2), 141-205.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and Transfer: A General Role for Analogical Encoding. *Journal of Educational Psychology, 95*(2), 393-408.

Goethals, G. R. (2001). *Peer Effects, Gender, and Intellectual Performance among Students at a Highly Selective College: A Social Comparison of Abilities Analysis. Discussion Paper*.

Goethals, G. R. (2002). *Social Comparison and Peer Effects at an Elite College. Discussion Paper*.

Gouli, E., Gogoulou, A., & Grigoriadou, M. (2008). Supporting Self-, Peer-, and Collaborative-Assessment in E-Learning: The Case of PEer and Collaborative ASSessment Environment (PECASSE). *Journal of Interactive Learning Research, 19*(4), 615-647.

Graner, M. H. (1987). Revision workshops: An alternative to peer editing groups. *The English Journal, 76*(3), 40-45.

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics Volume 2 Reading, writing, and language learning* (pp. 176-240). New York: Cambridge University Press.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59-87). New York: Academic Press.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups. *Administrative Science Quarterly, 44*(4), 741-763.

Jones, E. (2008). Predicting performance in first-semester college basic writers: Revisiting the role of self-beliefs. *Contemporary Educational Psychology, 33*(2), 209-238.

Kaufman, J., & Schunn, C. (in press). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*.

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review, 95*(4), 492-527.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research, 66*(4), 423-458.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30-43.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., et al. (2000). Manual for the Patterns of Adaptive Learning Scales (PALS). University of Michigan.

Miller, G., & Polito, T. (1999). The Effect of Cooperative Learning Team Compositions on Selected Learner Outcomes. *Journal of Agricultural Education, 40*(1), 66-73.

Moshman, M., & Geil, D. (1998). Collaborative Reasoning: Evidence for Collective Rationality. *Thinking & Reasoning, 4*(3), 231-248.

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375-401.

Newell, A. (1994). *Unified theories of cognition*: Harvard University Press.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Earlbaum.

Novick, L. R. (1988). Analogical Transfer, Problem Similarity, and Expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 510-520.

Ohlsson, S. (2009). Resubsumption: A Possible Mechanism for Conceptual Change and Belief Revision. *Educational Psychologist, 44*(1), 20-40.

Patchan, M. M., Charney, D., & Schunn, C. D. (2009). A validation study of students' end comments: Comparing comments by students, a writing instructor, and a content instructor. *Journal of Writing Research, 1*(2), 124-152.

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (under review). The effects of skill diversity on commenting and revisions.

Sager, C. (1973). *Improving the Quality of Written Composition Through Pupil Use of Rating Scale*. Paper presented at the 63rd Annual Meeting of the National Council of Teachers of English.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Siegler, R. S., & Stern, E. (1998). Conscious and Unconscious Strategy Discoveries: A Microgenetic Analysis. *Journal of Experimental Psychology: General, 127*(4), 377-397.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631-645.

Topping, K. J., & Ehly, S. (1998). *Peer-assisted learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*, 1161-1174.

Tutty, J. I., & Klein, J. D. (2008). Computer-mediated instruction: a comparison of online and face-to-face collaboration. *Educational Technology Research and Development, 56*(2), 101-124.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge: Harvard University Press.

Wallace, D. L., & Hayes, J. R. (1991). Redefining Revision for Freshmen. *Research in the Teaching of English, 25*(1), 54-66.

Watson, S. B., & Marshall, J. E. (1995). Effects of Cooperative Incentives and Heterogeneous Arrangement on Achievement and Interaction of Cooperative Learning Groups in a College Life-Science Course. [Article]. *Journal of Research in Science Teaching, 32*(3), 291-299.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive Activity and Learning in Collaborative Small Groups. *Journal of Educational Psychology, 87*(3), 406-423.

Wooley, R. S., Was, C., Schunn, C. D., & Dalton, D. (2008). *The effects of feedback elaboration on the giver of feedback*. Paper presented at the Cognitive Science.

6.0 APPENDIX A

STUDENT HANDOUTS

6.1. ASSIGNMENT: PAPER 1 DRAFT 1

Due Date: Submit your reviews on turnitin.com by 5pm on Friday, September 18.

Overview. Many psychological studies are commonly discussed in the media. For example, a recent report focused on the results of a study about playing video games that will be published in the American Journal of Preventative Medicine. It is our responsibility to critically evaluate the conclusions being drawn to decide whether we agree with them. The goal of this paper is to practice critical thinking and summarizing research articles. Your paper should be 3 pages in length (double-spaced). Your peers will provide feedback about how to improve your first draft. Your TA will provide a grade for both drafts. The paper submission and feedback will be done online. The peer review will be completely anonymous.

Step 1. Read the MSNBC article and the original research report. These articles can be found on Blackboard in the writing assignment folder.

- Choney, S. (2009). Study: Average gamer is 35, fat and bummed—CDC study finds playing leads to 'lower extraversion' in adult gamers. http://www.msnbc.msn.com/id/32463904/ns/technology_and_science-games
- Weaver, J. B., Mayes, D. Weaver, S. S., Kannenberg, W., Hopkins, G. L., Eroglu, E., & Bernhardt, J. M. (2009). Health-risk correlates of video-game playing among adults. *American Journal of Preventive Medicine.*

Step 2. Summarize both articles. When summarizing the research article, make sure to identify the hypotheses, independent variable(s) and the levels, dependent variable(s), the design of each study, and the outcome. You will need to explain how the outcome of this study supports your conclusions.

Step 3. Evaluate whether the MSNBC article provided an accurate report of the research study. Describe why causal conclusions cannot be made from correlational studies. Provide an alternative explanation about why playing video games is associated with poor health.

Step 4. Start writing. Flesh out the outline.

- Make it clear. Make sure your main argument is clearly identified. Make sure each paragraph has a (single) clear point. Make sure the ordering of points from paragraph to paragraph makes sense. Consider using section headers. Give concrete examples.
- Make it logical. You are trying to explain something to your peers. Help them understand why they should believe you. Provide evidence for claims. Just because you said so is not good enough.
- Make it interesting. You are trying to explain something to your peers that they don't already know. Make them feel like they learned something.
- Don't plagiarize (now or ever!). Don't quote large sections of other text as proof—very boring, even if properly cited. Re-describe things in your own words—make sure to cite where you got evidence if you are using particular sources of evidence.

Step 5. Clean up the paper one more time. Let the paper sit for a couple of days. Print it out. Reread it. Ask yourself: Does the prose flow well? Is it logical? Is it insightful? How can I make it better? If you having trouble writing in general, consider visiting the Writing Center located in LIB 206, Tutoring and Learning Services (stop in or call 813-974-2713 for an appointment). They are generally very helpful with these sorts of tasks.

Step 6. Submit your paper on turnitin.com. Make sure you do not include your name anywhere in your document, so that the peer-review process is anonymous. If you have any problems, immediately contact your TA or Melissa Patchan (melissa.patchan@gmail.com).

Step 7. Complete the Writing 1 survey on Blackboard by 5pm on Monday, September 21—survey will become available after Friday, September 18 after 5pm.

6.2. ASSIGNMENT: REVIEW 4 PEERS' PAPER 1 DRAFT 1

Due Date: Submit your reviews on turnitin.com by 5pm on Friday, September 25.

Overview. You will now review four peers' papers. Reviewing involves providing constructive feedback and evaluating the quality of the paper. Peer-review can benefit you in several ways.

First, you can receive a significant amount of feedback from peers (more than your single TA could provide), which could help you to revise drafts more successfully. Second, you will learn how to evaluate writing by providing feedback to your peers, which can help improve your general writing skills. Your TA will look over your reviews to make sure that you took your review task seriously.

Other things to consider:

- Be nice:
  - Mention the strengths of the paper, so people know where the succeeded
  - Flaming is not helpful to anyone (and will lead to a bad reviewing grade)
  - Remember that some people are starting off as worse writers than others. You are trying to help them improve their writing, not evaluate them as a person.

- Be open-minded about style:
  - There is no one fixed way these papers have to be written. For example, the paper doesn't have to be formal or informal.
  - Ask yourself: did you understand the paper, did you believe the argument, did you learning something?
  - If you say yes to all three, then the paper should be evaluated as good.

- Be constructive
  - Give particular ideas for how to improve the paper
  - Don't just complain about a problem; give some ideas for how to fix it

- Be specific
  - Be precise about where particular problems occur
  - Give exact examples of problems

- Think beyond yourself
  - You are trying to give advice on how to make this paper better for a broad audience of peers, not just you!

Step 1. Access the papers assigned to you. These will be available on turnitin.com by midnight on Saturday, September 19.

Step 2. The first step in reviewing your peers' papers is to provide comments that will help them revise their paper. You will be commenting on 3 dimensions: flow, argument logic, and insight. You should type your comments in a Word document first, and then go to turnitin.com and cut and paste your comments into the form.

- When structuring your comments, try to separate different ideas by using bullets, numbering, or separate lines for each idea. Arrange your comments with the most important issues near the top of the list.
- Be sure to give specific advice for how to fix any problems.
- Also, make sure to include some positive comments. If you thought some aspect of these sections was done well, mention what was good about it.
- Note that none of the dimensions consider low-level writing problems, like typos and simple grammar problems. Please limit these types of comments to issues that are so bad that they make it hard to understand the text.

Consider the following questions in order to provide constructive feedback that your peers could use to improve their paper.

Flow. Did the writing flow smoothly so you could follow the main argument? Did you understand what each of the arguments was and the ordering of the points made sense to you? First summarize the main points of the paper. Then make specific comments about problems you had in understanding the arguments and following the flow across arguments.

Argument Logic. 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 were they irrelevant to the claim being made or contradictory to the claim being made? Did the author consider obvious counter-arguments, or were they just ignored? If points were made without support, describe which ones there were. If some obvious counter-argument was not considered, explain what it was. Did the author clearly and accurately describe why causal conclusions cannot be made from correlational studies? Did the author accurately identify the hypothesis, independent variable(s) and the levels, dependent variable(s), the design of each study, and the outcome of the study being summarized?

Insight. 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? First summarize the main insights of the paper. What did you learn if anything? Did the author provide an original and interesting alternative explanation?

Step 3. The second step in reviewing your peers' papers is to rate various aspects of the paper. The ratings will range from 1 to 5 with a rating of 1 meaning 'Very Poor' and a rating of 5 meaning 'Very Good'.

On a scale of 1 to 5, rate…

Flow

…how well this paper stayed on topic.

1. It was unclear what the main topic of the paper was.
2. The paper rarely stayed on the main topic.
3. The paper sometimes stayed on topic.
4. The paper often stayed on topic.
5. The paper always or almost always stayed on topic.

…how well this paper was organized.

1. The paper was not organized at all.
2. The paper was slightly organized.
3. The paper was moderately organized.
4. The paper was very organized.
5. The paper was extremely organized.

Argument Logic

…how persuasively the paper makes its case.

1. The paper was not persuasive at all.
2. The paper was slightly persuasive.
3. The paper was moderately persuasive.
4. The paper was very persuasive.
5. The paper was extremely persuasive.

…how well the author explained why causal conclusions cannot be made from correlational studies.

1. The author did not identify nor explain why causal conclusions cannot be made from correlational studies.
2. The author incorrectly identified why causal conclusions cannot be made from correlational studies.
3. The author correctly identified but did not provide an explanation about why causal conclusions cannot be made from correlational studies.
4. The author correctly identified but inaccurately explained why causal conclusions cannot be made from correlational studies.
5. The author correctly identified and correctly explained why causal conclusions cannot be made from correlational studies.

…whether all of the relevant information from the research article was provided.

1. There was no summary of an article.
2. The summary included little of the relevant information.
3. The summary included some of the relevant information.
4. The summary included most of the relevant information.
5. The summary included all of the relevant information.

Insight

…how interesting and original this paper's conclusion was to you.

1. The paper was not interesting or original at all.
2. The paper was slightly interesting and original.
3. The paper was moderately interesting and original.
4. The paper was very interesting and original.
5. The paper was extremely interesting and original.

Step 4. Submit your reviews on turnitin.com. Turnitin.com does not include all of the above details, so make sure to use the guidelines provided above. If you have any problems, immediately contact your TA or Melissa Patchan (melissa.patchan@gmail.com).

Step 5. Complete the Reviewing 2 survey on Blackboard by 5pm on Monday, September 28—survey will become available after 5pm on Friday, September 25.

6.3. ASSIGNMENT: PAPER 1 DRAFT 2

Due Date: Submit your reviews on turnitin.com by 5pm on Friday, October 2.

Overview. For this paper, you were to critically evaluate whether the media accurately reported a recent psychological study. The goal of this paper is to practice critical thinking and summarizing research articles. Your paper should be 3 pages in length (double-spaced). Your peers have provided you feedback about how to improve your first draft. Your task for this assignment is to revise your paper based on that feedback. Your TA will provide a grade for both drafts. The paper submission and feedback will be done online. The peer review will be completely anonymous.

Step 1. Access your feedback on turnitin.com. The feedback will become available after 5pm on Friday, September 25.

- Read through all of the reviews.
- Make a list of the things that you will need to change. Try to address all of the issues brought up by your peers. Even if you do not agree with the feedback, your peer was confused by something, so you should try to make your paper clearer.

Step 2. Revise your draft.

- Remember to make it clear, make it logical, and make it interesting. See handout for draft 1 for more details.
- Don't plagiarize (now or ever!). Don't quote large sections of other text as proof—very boring, even if properly cited. Re-describe things in your own words—make sure to cite where you got evidence if you are using particular sources of evidence.

Step 5. Clean up the paper one more time. Let the paper sit for a couple of days. Print it out. Reread it. Ask yourself: Does the prose flow well? Is it logical? Is it insightful? How can I make it better? If you having trouble writing in general, consider visiting the Writing Center located in LIB 206, Tutoring and Learning Services (stop in or call 813-974-2713 for an appointment). They are generally very helpful with these sorts of tasks.

Step 6. Submit your paper on turnitin.com. If you have any problems, immediately contact your TA or Melissa Patchan (melissa.patchan@gmail.com).

Step 7. Complete the Writing 4 survey on Blackboard by 5pm on Monday, October 5—survey will become available after 5pm on Friday, October 2.

6.4. ASSIGNMENT: PAPER 2

Due Date: Submit your paper on turnitin.com by 5pm on Monday, November 23.

Overview. Do you know someone who has met his or her partner or spouse through an Internet matching service? Computer dating services have been around for a very long time. As far back as the 1970s, people began to think that by using the power of the electronic brain, they could find the one person destined for them (Lum & Curran, 1975). Originally, these services matched people based on answers to basic questions about their likes and dislikes, lifestyle preferences, religion, and other such variables. As computers became more powerful, the programs became more sophisticated. Several companies offering matchmaking services on the Internet use personality tests as the basis for their pairings (Carter, 2005; Hansell, 2004). One company even claims to help people find their "soul mates," although as Gottlieb (2005) reports, that may be more complicated than some people anticipate. You can find information about some of these Internet based companies by looking at their websites (e.g., http://eharmony.com/servlet/about/eharmony or http://www.match.com/help/aboutus.aspx).

As you might imagine, anything to do with love is going to attract a lot of attention. Many researchers have looked at how people are attracted to one another, the different kinds of love that people feel, and what makes relationships last (Gattis, Berns, Simpson, & Christensen, 2004;

Murstein & Brust, 1985; Sternberg, 2000; Zentner, 2005). Others have investigated the expectations of people who use computer-matching services (Houran & Lange, 2004).

For your second paper, use the critical thinking method to come up with your own conclusion about whether Internet matchmaking sites are likely to provide successful results. The goal of this paper is to practice critical thinking and summarizing research articles. Your paper should be 3 pages in length (double-spaced).

Step 1. Use the critical thinking method to come to your own conclusions about whether Internet matchmaking sites are likely to provide successful results. Be sure to support your thoughts with credible source material. Remember that in thinking critically, you need to answer the following questions:

- What am I being asked to believe?
- What evidence is there to support the assertion?
- Are there alternative ways of interpreting the evidence?
- What additional evidence would help to evaluate the alternatives?
- What conclusions are most reasonable?

Step 2. You should provide some in-depth evidence for your conclusions by focusing on the details of one of the articles cited in the overview. Read one of the following articles. All of these articles can be found on Blackboard.

- Gattis, K. S., Bern, S., Simpson, L. E., & Christensen, A. (2004). Birds of a feather or strange birds? Ties among personality dimensions, similarity, and marital quality. *Journal of Family Psychology, 18(4),* 564-574.
- Houran, J. & Lange, R. (2004). Expectations of finding a 'soul mate' with online dating. *North American Journal of Psychology, 6(2),* 297-308.
- Lum., K. & Curran, J. P. (1975). Personality similarity and interpersonal attraction in the computer dating situation. *Journal of Social Psychology, 95(2),* 233-239.
- Murstein, B. I. & Brust, R. G. (1985). Humor and interpersonal attraction. *Journal of Personality Assessment, 49(6),* 637-640.
- Zentner, M. R. (2005). Ideal mate personality concepts and compatibility in close relationships: A longitudinal analysis. *Journal of Personality and Social Psychology, 89(2),* 242-256.

Step 3. Summarize the article. Make sure to identify the hypotheses, independent variable(s) and the levels, dependent variable(s), the design of the study, and the outcome. You will need to explain how the outcome of this study supports your conclusions.

Step 4. Start writing. Flesh out the outline.

- Make it clear. Make sure your main argument is clearly identified. Make sure each paragraph has a (single) clear point. Make sure the ordering of points from paragraph to paragraph makes sense. Consider using section headers. Give concrete examples.

- Make it logical. You are trying to explain something. Help the reader understand why he/she should believe you. Provide evidence for claims. Just because you said so is not good enough.
- Make it interesting. You are trying to explain something to your readers. Make them feel like they learned something.
- Don't plagiarize (now or ever!). Don't quote large sections of other text as proof—very boring, even if properly cited. Re-describe things in your own words—make sure to cite where you got evidence if you are using particular sources of evidence.

Step 5. Clean up the paper one more time. Let the paper sit for a couple of days. Print it out. Reread it. Ask yourself: Does the prose flow well? Is it logical? Is it insightful? How can I make it better? If you having trouble writing in general, consider visiting the Writing Center located in LIB 206, Tutoring and Learning Services (stop in or call 813-974-2713 for an appointment). They are generally very helpful with these sorts of tasks.

Step 6. Submit your paper on turnitin.com. If you have any problems, immediately contact your TA or Melissa Patchan (melissa.patchan@gmail.com).

Step 7. Complete the Writing 3 survey on Blackboard by 5pm on Friday, November 27—survey will become available on Monday, November 23 after 5pm.

7.0 APPENDIX B

WRITING QUALITY RUBRICS

Flow

…how well the paragraphs were developed.

5 – Very Good	All paragraphs stated a point and developed it
4 – Good	Most paragraphs stated a point and developed it
3 – Fair	Some paragraphs stated a point and developed it. All paragraphs introduced a topic, but may not state an explicit point
2 – Poor	Some paragraphs stated a point OR introduced a topic, but did not develop it
1 – Unsatisfactory	No paragraphs stated a point and/or paragraphs shifted topics frequently.

…how well transitions connected paragraphs.

5 – Very Good	Strong transitions between all paragraphs
4 – Good	Strong transitions between most paragraphs
3 – Fair	Transitions between most paragraphs, but some were weak
2 – Poor	Weak transitions between some of the paragraphs
1 – Unsatisfactory	No transitions between paragraphs

…how well this paper was organized around a main idea.

5 – Very Good	All paragraphs were connected to the main point
4 – Good	Most paragraphs were connected to the main point
3 – Fair	Some paragraphs were connected to the main point
2 – Poor	Most paragraphs were not connected to the main point
1 – Unsatisfactory	No main point explicitly stated

Argument Logic (Paper 1)

…how well the author evaluated the MSNBC article.

5 – Very Good	All points were supported by concrete evidence or examples
4 – Good	Most points were supported by concrete evidence or examples
3 – Fair	Some points were supported by concrete evidence or examples
2 – Poor	Few points were supported by concrete evidence or examples
1 – Unsatisfactory	No support was provided

…how well the author explained causal conclusions.

5 – Very Good	Provided a complete and clear explanation (i.e., $A \rightarrow B$; $B \leftarrow A$; $C \rightarrow [A \rightarrow B]$)
4 – Good	Provided a complete and somewhat clear explanation
3 – Fair	Provided complete but unclear explanation
2 – Poor	Provided an incomplete explanation
1 – Unsatisfactory	No explanation was provided

…how well the author explained an alternative possibility.

5 – Very Good	Provided an appropriate and clear alternative
4 – Good	Provided an appropriate and somewhat clear alternative
3 – Fair	Provided an appropriate alternative, but did not explain it
2 – Poor	Provided an inappropriate alternative
1 – Unsatisfactory	No alternative possibility was provided

…whether all the required information from the research article was accurately provided.

5 – Very Good	The summary accurately included all of the required information
4 – Good	The summary accurately included most of the required information
3 – Fair	The summary accurately included some required information
2 – Poor	The summary included little required information OR the information was inaccurate
1 – Unsatisfactory	No summary of the article

Argument Logic (Paper 2)

…how appropriate was the evidence provided.

5 – Very Good	All points are supported by concrete evidence or examples
4 – Good	Most points were supported by concrete evidence or examples
3 – Fair	Some points were supported by concrete evidence or examples
2 – Poor	Few points were supported by concrete evidence or examples
1 – Unsatisfactory	No support was provided

…how relevant were the conclusions.

5 – Very Good	All conclusions were relevant
4 – Good	Most conclusions were relevant
3 – Fair	Some conclusions were relevant
2 – Poor	Most conclusions were not relevant
1 – Unsatisfactory	No conclusions were offered

Insight

…how well the main point was connected to a larger issue.

5 – Very Good	Main point was fully connected to a relevant larger issue throughout the whole paper
4 – Good	Main point was connected to a relevant larger issue
3 – Fair	Some points demonstrated an innovative analysis, but these points were not connected to a relevant larger issue
2 – Poor	One point demonstrated an innovative analysis, but this point was not connected to a relevant larger issue
1 – Unsatisfactory	No points demonstrated an innovative analysis

8.0 APPENDIX C

DESCRIPTIVE STATISTICS FOR STUDY 1

Table 9. Descriptive Statistics: Revision Quality (Study 1)

High Writer

	High Feedback Source										
	Comment Type Present				Comment Type Absent						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	149	0.22	0.48	0.04	85	0.29	0.60	0.07	-0.13	0.9	.34
solution	132	0.22	0.48	0.04	102	0.29	0.62	0.06	-0.13	1.0	.33
localization	75	0.16	0.49	0.06	159	0.36	0.45	0.04	-0.42	2.9	< .01
high prose	89	0.38	0.59	0.06	145	0.14	0.51	0.04	0.44	3.2	< .01
substance	103	0.38	0.64	0.06	131	0.14	0.48	0.04	0.42	3.1	< .01
	Low Feedback Source										
	Comment Type Present				Comment Type Absent						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	122	0.19	0.47	0.04	44	0.34	0.52	0.08	-0.31	1.7	.09
solution	94	0.22	0.46	0.05	72	0.31	0.61	0.07	-0.16	1.0	.32
localization	65	0.24	0.53	0.07	101	0.30	0.48	0.05	-0.12	0.7	.46
high prose	84	0.34	0.64	0.07	82	0.19	0.50	0.06	0.28	1.8	.08
substance	50	0.41	0.55	0.08	116	0.12	0.54	0.05	0.52	3.1	< .01

Low Writer

	High Feedback Source										
	Comment Type Present				Comment Type Absent						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	173	0.18	0.49	0.04	79	0.18	0.52	0.06	0.01	0.1	.93
solution	147	0.15	0.50	0.04	105	0.21	0.56	0.06	-0.11	0.8	.42
localization	121	0.10	0.54	0.05	131	0.26	0.46	0.04	-0.32	2.5	.01
high prose	93	0.24	0.56	0.06	159	0.13	0.48	0.04	0.21	1.6	.11
substance	88	0.27	0.57	0.06	164	0.09	0.47	0.04	0.33	2.4	.02
	Low Feedback Source										
	Comment Type Present				Comment Type Absent						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	143	0.25	0.53	0.04	97	0.29	0.59	0.06	-0.06	0.5	.63
solution	138	0.29	0.51	0.04	102	0.25	0.61	0.06	0.08	0.6	.57
localization	72	0.20	0.48	0.06	168	0.34	0.45	0.04	-0.31	2.2	.03
high prose	102	0.35	0.58	0.06	138	0.19	0.49	0.04	0.30	2.3	.02
substance	87	0.44	0.57	0.06	153	0.10	0.48	0.04	0.63	4.6	< .001

Table 10. Descriptive Statistics: Amount & Type of Comments (Study 1)

	High Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	47	9.8	5.5	0.8	46	9.5	5.4	0.8	0.05	0.3	.79
solution	47	9.4	4.8	0.7	46	7.2	4.7	0.7	0.46	2.2	.03
localization	47	6.1	4.1	0.6	46	5.9	4.1	0.6	0.05	0.2	.81
high prose	47	8.5	4.8	0.7	46	7.5	4.7	0.7	0.21	1.0	.32
substance	47	5.6	3.4	0.5	46	4.2	3.4	0.5	0.41	2.0	.05
	Low Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	49	12.7	5.6	0.8	47	9.2	5.5	0.8	0.63	3.1	< .01
solution	49	9.9	4.9	0.7	47	9.3	4.8	0.7	0.12	0.6	.55
localization	49	7.9	4.2	0.6	47	5.9	4.1	0.6	0.48	2.4	.02
high prose	49	9.3	4.2	0.6	47	7.9	4.8	0.7	0.31	1.5	.13
substance	49	6.4	3.5	0.5	47	4.5	3.4	0.5	0.55	2.7	.01

Table 11. Descriptive Statistics: Implementability (Study 1)

	High Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	395	0.42	0.56	0.03	366	0.36	0.55	0.03	0.09	1.3	.20
solution	412	0.38	0.59	0.03	305	0.33	0.52	0.03	0.08	1.1	.28
localization	172	0.44	0.52	0.04	164	0.34	0.54	0.04	0.18	1.6	.11
high prose	325	0.33	0.70	0.04	278	0.26	0.73	0.04	0.09	1.1	.25
substance	252	0.40	0.67	0.04	186	0.25	0.65	0.05	0.23	2.4	.02
	Low Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
problem	548	0.35	0.56	0.02	376	0.42	0.58	0.03	-0.13	-2.0	.05
solution	450	0.33	0.53	0.03	379	0.40	0.58	0.03	-0.13	-1.8	.07
localization	264	0.41	0.55	0.03	149	0.52	0.51	0.04	-0.21	-2.1	.04
high prose	369	0.22	0.69	0.04	303	0.41	0.71	0.04	-0.28	-3.6	< .001
substance	312	0.25	0.67	0.04	203	0.46	0.63	0.04	-0.33	-3.7	< .001

Table 12. Descriptive Statistics: Draft Quality (Study 1)

	High Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
Paper 1 Draft 1	47	2.3	0.5	0.1	46	2.4	0.5	0.1	-0.13	-0.6	.54
Paper 1 Draft 2	47	2.6	0.2	0.0	46	2.6	0.2	0.0	0.17	0.8	.40
Paper 2	35	2.5	0.5	0.1	29	2.6	0.5	0.1	-0.13	-0.5	.62
	Low Writer										
	High Feedback Source				Low Feedback Source						
	N	M	SD	SE	N	M	SD	SE	d	t	p
Paper 1 Draft 1	49	2.0	0.5	0.1	47	2.1	0.5	0.1	-0.07	-0.4	.71
Paper 1 Draft 2	49	2.2	0.2	0.0	47	2.3	0.2	0.0	-0.17	-0.8	.40
Paper 2	35	2.3	0.4	0.1	32	2.2	0.4	0.1	0.29	1.2	.24

9.0 APPENDIX D

INFERENTIAL STATISTICS FOR STUDY 1

Table 13. Inferential Statistics: Draft Quality (Study 1)

	Paper 1 Draft 1
writer ability	F(1, 181) = 22.7, p = .00
text quality	F(1, 181) = 3.5, p = .06
feedback source	F(1, 181) < 1, p = .49
writer ability * text quality	F(1, 181) = 2.1, p = .15
writer ability * feedback source	F(1, 181) < 1, p = .85
text quality * feedback source	F(1, 181) < 1, p = .95
writer ability * text quality * feedback source	F(1, 181) = 1.3, p = .26

	Paper 1 Draft 2 High Writer	Paper 1 Draft 2 Low Writer
paper 1 draft 1 (covariate)	F(1, 88) = 49.3, p = .00	F(1, 91) = 348.9, p = .00
text quality	F(1, 88) < 1, p = .73	F(1, 91) < 1, p = .73
feedback source	F(1, 88) < 1, p = .41	F(1, 91) < 1, p = .40
text quality * feedback source	F(1, 88) < 1, p = .63	F(1, 91) < 1, p = .85

	Paper 2 High Writer	Paper 2 Low Writer
paper 1 draft 1 (covariate)	F(1, 59) = 17.6, p = .00	F(1, 62) = 3.4, p = .07
text quality	F(1, 59) = 1.4, p = .25	F(1, 62) = 3.9, p = .05
feedback source	F(1, 59) < 1, p = .62	F(1, 62) = 1.4, p = .24
text quality * feedback source	F(1, 59) < 1, p = .61	F(1, 62) = 1.6, p = .21

Table 14. Inferential Statistics: Amount & Type of Comments (Study 1)

	writer ability	feedback source	interaction
problem	F(1, 185) = 2.5, p = .12	F(1, 185) = 5.5, p = .02	F(1,185) = 4.1, p = .05
solution	F(1, 185) = 3.5, p = .07	F(1, 185) = 4.3, p = .04	F(1, 185) = 1.3, p = .25
localization	F(1, 185) = 2.1, p = .15	F(1, 185) = 3.3, p = .07	F(1, 185) = 1.9, p = .17
high prose	F(1, 185) < 1, p = .36	F(1, 185) = 3.1, p = .08	F(1, 185) < 1, p = .72
substance	F(1, 185) = 1.4, p = .23	F(1, 185) = 10.8, p = .001	F(1, 185) < 1, p = .63

Table 15. Inferential Statistics: Implementability & Revision Quality (Study 1)

	Implementability	Revision Quality
writer ability	$F(1, 2690) = 3.4$, $p = .07$	$F(1, 868) < 1$, $p = .35$
feedback source	$F(1, 2690) = 1.2$, $p = .28$	$F(1, 868) = 1.7$, $p = .20$
problem	$F(1, 2690) = 8.2$, $p = .004$	$F(1, 868) = 2.7$, $p = .10$
solution	$F(1, 2690) < 1$, $p = .39$	$F(1, 868) = 1.3$, $p = .25$
localized	$F(1, 2690) = 53.0$, $p < .001$	$F(1, 868) = 2.3$, $p < .001$
high prose	$F(1, 2690) = 13.2$, $p < .001$	$F(1, 868) = 19.2$, $p < .001$
substance	$F(1, 2690) < 1$, $p = .45$	$F(1, 868) = 41.1$, $p < .001$
writer ability * feedback source	$F(1, 2690) = 2.6$, $p < .001$	$F(1, 868) = 1.2$, $p = .28$
writer ability * problem	$F(1, 2690) = 4.2$, $p = .04$	$F(1, 868) = 1.6$, $p = .21$
writer ability * solution	$F(1, 2690) = 2.2$, $p = .14$	$F(1, 868) < 1$, $p = .34$
writer ability * localized	$F(1, 2690) = 1.5$, $p = .23$	$F(1, 868) < 1$, $p = .72$
writer ability * high prose	$F(1, 2690) = 1.3$, $p = .26$	$F(1, 868) < 1$, $p = .40$
writer ability * substance	$F(1, 2690) < 1$, $p = .61$	$F(1, 868) < 1$, $p = .92$
feedback source * problem	$F(1, 2690) < 1$, $p = .54$	$F(1, 868) < 1$, $p = .44$
feedback source * solution	$F(1, 2690) < 1$, $p = .60$	$F(1, 868) < 1$, $p = .58$
feedback source * localized	$F(1, 2690) < 1$, $p = .39$	$F(1, 868) = 1.5$, $p = .22$
feedback source * high prose	$F(1, 2690) = 2.2$, $p = .14$	$F(1, 868) < 1$, $p = .82$
feedback source * substance	$F(1, 2690) < 1$, $p = .81$	$F(1, 868) = 1.6$, $p = .20$
writer ability * feedback source * problem	$F(1, 2690) = 3.8$, $p = .05$	$F(1, 868) < 1$, $p = .81$
writer ability * feedback source * solution	$F(1, 2690) = 4.7$, $p = .03$	$F(1, 868) < 1$, $p = .46$
writer ability * feedback source * localized	$F(1, 2690) < 1$, $p = .59$	$F(1, 868) = 1$, $p = .31$
writer ability * feedback source * high prose	$F(1, 2690) < 1$, $p = .47$	$F(1, 868) < 1$, $p = .38$
writer ability * feedback source * substance	$F(1, 2690) = 6.9$, $p = .01$	$F(1, 868) < 1$, $p = .48$

10.0 APPENDIX E

INFERENTIAL STATISTICS FOR STUDY 2

Table 16. Inferential Statistics: Amount, Features, and Focus of Comments (Study 2)

	word count provided	# of comments provided
P1D1Total	F(1, 181) = 4.7, p = .03	F(1, 181) < 1, p = .94
WriterAbility	F(1, 181) < 1, p = .79	F(1, 181) = 2.1, p = .15
Providing	F(1, 181) < 1, p = .53	F(1, 181) < 1, p = .34
WriterAbility * Providing	F(1, 181) = 1.0, p = .32	F(1, 181) < 1, p = .61

	# of problems provided	# of solutions provided	# of localized provided
P1D1Total	F(1, 181) = 12.3, p = .001	F(1, 181) = 2.4, p = .12	F(1, 181) = 3.6, p = .06
WriterAbility	F(1, 181) < 1, p = .71	F(1, 181) = 1.4, p = .24	F(1, 181) < 1, p = .95
Providing	F(1, 181) = 2.3, p = .13	F(1, 181) = 1.3, p = .26	F(1, 181) < 1, p = .34
WriterAbility * Providing	F(1, 181) = 6.3, p = .01	F(1, 181) < 1, p = .92	F(1, 181) = 1.6, p = .21

	# of high prose provided	# of low prose provided	# of substance provided
P1D1Total	F(1, 181) < 1, p = .84	F(1, 181) < 1, p = .41	F(1, 181) = 9.8, p = .002
WriterAbility	F(1, 181) = 5.1, p = .03	F(1, 181) < 1, p = .71	F(1, 181) = 1.3, p = .26
Providing	F(1, 181) = 3.4, p = .07	F(1, 181) < 1, p = .35	F(1, 181) < 1, p = .50
WriterAbility * Providing	F(1, 181) = 2.8, p = .10	F(1, 181) < 1, p = .65	F(1, 181) < 1, p = .37