
Commenting on Writing

Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts

Kwangsu Cho

Christian D. Schunn

University of Pittsburgh, Pittsburgh, PA

Davida Charney

University of Texas, Austin

How do comments on student writing from peers compare to those from subject-matter experts? This study examined the types of comments that reviewers produce as well as their perceived helpfulness. Comments on classmates' papers were collected from two undergraduate and one graduate-level psychology course. The undergraduate papers in one of the courses were also commented on by an independent psychology instructor experienced in providing feedback to students on similar writing tasks. The comments produced by students at both levels were shorter than the instructor's. The instructor's comments were predominantly directive and rarely summative. The undergraduate peers' comments were more mixed in type; directive and praise comments were the most frequent. Consistently, undergraduate peers found directive and praise comments helpful. The helpfulness of the directive comments was also endorsed by a writing expert.

Keywords: *peer feedback; peer reviews; writing in the disciplines; writing assessment; instructional technology*

As Richard Haswell (2005) noted in a recent article, peer review of undergraduate writing seems to be "the least studied of practices now very common in college writing classrooms" (p. 211). In fact, while conducting the first comprehensive review of research on any kind of college-level peer assessment, Keith Topping (1998) found only 67 empirical studies to include.

Authors' Note: This work was supported by grants from the Andrew Mellon Foundation and the Provost Office of the University of Pittsburgh to the second author. We thank Andrea Grana and Chioma Azi for their assistance in coding and their helpful advice on earlier versions of this article.

His summary of outcomes regarding peer review of writing is cautiously positive: peer review “appears capable of yielding outcomes at least as good as teacher assessment and sometimes better” (p. 262).

Even without a solid research base, however, peer review is ubiquitous in 1st-year composition. Composition instructors have come to see peer review as an essential practice, partly because it insures a round of drafting and revising and partly from an assumption that writers benefit both from commenting and from reading comments. However, given the myriad of ways in which peer review is enacted in writing classrooms, it is important to inquire further into the consequences, positive, and negative, of various aspects of the peer review process.

Many important questions remain unanswered (Boud, Cohen, & Sampson, 1999; Topping, 1998). Most previous studies have focused on numeric ratings assigned by student peers to analyze their validity and reliability (e.g., Cho & Schunn, 2003; Hughes & Large, 1993; Magin & Helmore, 2000; Mowl & Pain, 1995). However valid and reliable they may be, the value of peer comments for student writers is unknown. Anecdotal evidence suggests that students actually find the task of reading and commenting on peers’ papers to be more helpful for revising than attempting to address their peers’ suggestions. Students have reported concerns that peers don’t take the task seriously, aren’t as qualified as the instructor in the subject matter, have had too little training in writing or practice at making comments, and are simply not the readers assigning grades (Artemeva & Logie, 2002). Perhaps, then, producing comments is what is valuable in and of itself (Graner, 1985). However, the possibility that writers gain nothing from reading their peers’ comments seems remote. Students are capable of using instructors’ comments to make some improvements to their drafts; if students’ concerns about the quality of peers’ comments were allayed, they may well achieve similar improvements from peer comments. The research question of greatest interest, then, is the intrinsic quality of peer comments and their helpfulness to writers.

This question has been difficult to address in standard writing classes for two reasons. First, the effects of peer comments are difficult to isolate from the many alternative forms of support available to writers: from their instructor, from a writing center, and from their composition textbook. Second, the small number of students in any writing classroom will not produce a representative sampling of commenting possibilities.

Isolating the effects of peer comments may be best achieved outside the 1st-year writing classroom, in content courses with intensive writing components, such as in writing in the disciplines (WID) programs. A large

undergraduate class provides the numbers of participants needed in a context that is uniform in the instructor's abilities and the writing resources available. The daunting logistical task of conducting peer reviews in large classes has recently been rendered much easier by the development of online environments for exchanging papers, assigning evaluators, and delivering the reviews (e.g., TurnItIn.com, InSite.wadsworth.com). One such environment, SWORD (Cho & Schunn, in press) has been used at various universities since Fall 2002 to support writing and peer review in 20 sections of psychology, history, education, English composition, information science, and sports.

In this article, we report on a study of the comments that peers generate using SWORD in three psychology courses, two undergraduate classes, and one graduate class. We investigate the nature of the comments produced for expository and argumentative papers, focusing specifically on these questions:

- What types of comments do undergraduates and graduate students provide to their peers?
- How do undergraduates' comments compare to those of an instructor who is a subject-matter expert?
- What types of comments do undergraduate students perceive as helpful in revising their writing?

Types of Peer Comments

Several typologies of comments on writing have been proposed (Beason, 1993; Coupe, 1986; Clare, Valdes, & Patthey-Chavez, 2000), based on the scope of the comment (local vs. global), the topic (mechanics, organization, content), and the function (evaluation, suggestion, response). The topic dimension is important to investigate as an indication of a peer's ability and readiness to detect writing problems of different kinds. The function dimension is also of interest as an indication of the range of roles vis-à-vis the writer that peers can adopt while reading (coach, judge, teammate).

Writing instructors are known to adopt a variety of roles, acting sometimes as a judge, sometimes as a coach, and sometimes as a typical reader (Dragga, 1992; Fife & O'Neill, 2001). Not surprisingly, the evaluative role is usually the most frequent. Summer Smith (1997) coded instructors' end comments on 313 papers from 1st-year undergraduate composition courses. Of nearly 1,400 segments, 72% were coded as judging, 20% as coaching, and 8% as reader response. Smith also found that instructors typically alternated between praise and criticism, with certain topics (such as effort or interest) more likely to be praised.

Evaluative comments can provide praise or criticism, to which writers may well respond differently. Beyond simple complimenting and fault-finding, evaluative comments take on a positive or negative connotation depending on how they are stated. Comments that are mitigated with qualifiers and reasons are more likely to be seen as positive.

In a study of commenting among computer scientists, Neuwirth, Chandook, Charney, Wojahn, and Kim (1994) employed a technology that permitted reviewers to annotate a text with written or spoken comments. Speaking increased the fluency of the comments, facilitating the inclusion of more mitigating language. The researchers found that writers responded more favorably to peers who made spoken comments (with the more mitigating language), rating them as having more integrity and being more likeable. These results are consistent with studies of students' reactions to instructor comments (Stevens, 1973; Taylor & Hoedt, 1966).

Praise and criticism have also been found to influence a writer's responses. Gee (1972) found that high school juniors who never received praise turned in shorter final drafts. Seidman (1968) found that high school students who consistently received positive feedback across eight writing assignments produced more optional rough drafts and revisions than either those who received negative comments or no comments. Consistently, some social psychologists (Butler, 1987; Shrauger & Schoenemann, 1979) have argued that positive evaluation information is more likely to be accepted than negative information.

The specificity and directiveness of comments are also important factors. Students are less likely to improve their writing if instructors' comments are vague or ambiguous (Beason, 1993; Clare et al., 2000; Coupe, 1986). On the other hand, very directed comments may lead to changes only in the specific draft and not lead to general changes in writing behavior.

Peer and Instructor Comments

A second focus of this study is to compare the nature of comments generated by student peers and subject-matter experts who are typical instructors in subject-matter courses. Assumptions about the great value of comments from disciplinary experts underlie the development of WID programs. In fact, giving students repeated opportunities to practice writing within a disciplinary context has produced some positive results. Recently, Johnstone, Ashbaugh, and Warfield (2002) found that students who were enrolled in an integrated multiyear writing initiative in accounting improved their writing during several semesters more than other business students

who had only taken a general business writing course with sporadic writing assignments in other courses.

However, the nature of comments from subject-matter instructors has not been carefully studied, especially the differences between their comments and those of writing instructors. Subject-matter experts may be more capable of commenting on the content than on the writing. Summer Smith (2003) compared think-aloud comments on a set of four engineering proposals from six technical writing instructors and six engineering faculty members. She found that the instructors produced identical rankings of the proposals but differed in the topic of the comments and their content. Engineering faculty frequently judged whether students' claims were valid and mentioned relevant information from their own experience, whereas writing instructors produced far more judgments of the rhetorical appropriateness of the writing with respect to the intended audience.

Studies comparing the nature of peer and instructor comments are rare. On its face, students might be expected to produce less helpful comments because they know less about the subject matter, are less skilled at writing, and have less experience evaluating papers than subject-matter instructors. There is some evidence, however, that peer comments may be more helpful in some ways than those of instructors. Richer (1992) found that oral discussions of student papers by peers led to marginally better writing on a posttest than discussions by instructors.

Cho and Schunn (in press) investigated expert-perceived improvement in revised drafts after undergraduate writers received comments from a subject-matter expert or from one or more peers. The results suggested that feedback from multiple peers may be more effective than from a single subject-matter expert.

Perceived Helpfulness

A final important factor in assessing the quality of comments is how helpful writers perceive them to be. Perceived helpfulness is likely to mediate between the feedback and the revisions made in later writing (Rucker & Thompson, 2003). Consistent with this view, Atwater, Waldman, Atwater, and Cartier (2000) found that feedback was less effective for those who had negative perception about the feedback process than those with positive perception. So, if students perceive peer feedback as less helpful than instructor feedback, this perception difference per se could be important in influencing revision behavior. Similarly, it is important to know what types of feedback students perceive to be helpful. For example, if students do not

perceive criticism comments to be helpful, that perception per se may influence the value of critical comment for draft revisions.

Method

Overview

In this study, we examine the types of comments that reviewers produce, coded by topic and by function, as well as their perceived helpfulness. Comments on classmates' papers were collected from two undergraduate-level courses and one graduate-level psychology course. The undergraduate papers in one class were also commented on by an independent psychology instructor experienced in assigning similar writing tasks. This particular source of comments from undergraduates and a subject-matter expert is particularly useful for making the comparison because the papers for which comments were generated were exactly the same ones. The second undergraduate class was selected to be a rather different type of class from the first, to assess the generality of findings regarding the nature of undergraduate feedback.

The graduate student course was selected as a naturalistic comparison against the undergraduate course data; the comparison is not one specifically of developing expertise situated between undergraduates and a subject-matter expert but rather more generally regarding how graduate-student-on-graduate-student peer comments compare to undergraduate-student-on-undergraduate-student comments. Graduate students are also often asked to provide peer feedback in their coursework, and it is important to find out how they tend to approach this task. Graduate students have received considerable training in critical thinking in their own discipline and, thus, may be especially prone to providing overly critical feedback to their peers.

Participants and Setting

Small undergraduate course (UNDGRD-SM). The participants were 30 students enrolled in a 12-week class entitled "Research Methods" at the University of Pittsburgh. The course focused on skills such as sampling, control, experimental design, data collection, data analysis, and research writing. Two students dropped the course after writing their first draft. The remaining 28 students had an average of 3.4 years of college and included 7 males (25%) and 21 females (75%).

Subject matter expert. A subject-matter expert was hired to generate comments for the 28 papers produced in the UNDGRD-SM class, was given the same feedback prompt as the students but was also told to generate the amount of feedback that she normally would as the instructor in this kind of WID course. The subject-matter expert had a PhD in developmental psychology covering the subject matter of the writing assignment. She was not the instructor of the class but had taught similar college courses involving writing tasks on similar topics for the past 8 years. According to student-teaching evaluation records, her teaching was ranked within the top 35% (i.e., a good but not unusually good instructor). As a reliability check, her ratings of the paper quality were compared to those of another independent rater who directed the local writing center, reaching moderately high reliability, $r = .60, p < .05$. In addition, the subject-matter expert reevaluated a random selection of 30% of the student papers about 3 months after the course, achieving a high test-retest reliability, $r = .90, p < .01$.

Large undergraduate course (UNDGRD-LG). Eighty-eight student participants were registered in “Cognitive Psychology for Non-Majors” at the University of Pittsburgh. Of the 88 in the study, 28 were male (32%) and 60, were female (68%). Unlike the UNDGRD-SM, there was no expert or instructor feedback on any of the students’ writing. There was no experimental manipulation within this course.

Graduate course (GRAD). The participants were 23 graduate students enrolled in an introductory-level graduate course titled “Problem Solving and Reasoning” at the University of Pittsburgh. The students, 10 males (43%) and 13 females (57%), were pursuing studies in psychology or education. There was no experimental manipulation in this course nor was there expert feedback on the papers.

Procedure

In each course, individual students played two roles: as writer and as reviewer. Midway through the term, students followed a sequence of activities, with 1 week devoted to each of the following steps:

- [Author] writing a draft
- [Reviewer] commenting and evaluating on the drafts of six other students
- [Author] receiving comments, revising the draft, and rating the helpfulness of the reviewer comments
- [Reviewer] commenting and evaluating on the revised drafts of the same six students

Writing task. In each class, students chose a writing topic. (See Appendix A for writing prompts.) In two of the courses, UNDGRD-LG and GRAD, students wrote expository essays, choosing from a list of topic options. In the third course, UNDGRD-SM, students wrote the introductory section of a report on a research project that they were designing and conducting. All students were informed that they would receive ratings and comments on a first draft, revise that draft, and receive ratings and comments on a second draft. Students in UNDGRD-LG and GRAD were told that classmates would provide the ratings and comments. In UNDGRD-SM, where some comments were produced by the subject-matter expert, students were also told that reviewers would provide ratings and comments but that the reviewers would either be six peers, a single peer, or a single expert. They were also told that the scores would be balanced if systematic differences in ratings were found across the three feedback conditions.

Helpfulness ratings of reviewers' comments. After submitting their final drafts, students evaluated how helpful they found the reviewers' comments for revising. Helpfulness was evaluated on a 5-point rating scale from *not helpful at all* (1) to *very helpful* (5) with space provided for optional short responses. The helpfulness ratings were used to determine 50% of the reviewing grades; thus, students were given a good incentive to provide helpful comments. The other 50% of the reviewing grades were computed by the SWORD system algorithms on reviewer evaluation accuracy (see Cho & Schunn, in press, for details).

Reviewing task. Reviewers were given 1 week to read drafts and provide feedback. Each student reviewed the drafts of six classmates. In addition to the student reviewers, the subject-matter expert in UNDGRD-SM was given 1 week to rate and comment on all 28 drafts of both first and final versions, following the same instructions provided to the peer reviewers. The expert was not informed who would receive her reviews. Reviewers were blind to authors' identities as were authors to reviewers' identities.

Reviewers were required to evaluate each draft both qualitatively and quantitatively along three dimensions: prose flow, argument, and insight. For each dimension, they wrote comments and then provided a rating along a 7-point rating scale from *disastrous* (1) to *excellent* (7). Prose flow concerned how easily the main points of the argument could be followed, including aspects of sequencing and transitions. Argument concerned the quality of the claims and support, including relevance and consideration of counter-arguments. Insight concerned the new ideas, information, and

inferences that the paper contributed to the class, beyond assigned course texts and materials. See Appendix B for instructions to reviewers about these dimensions and Appendix C for a screen shot of the instructions on the SWORD system.

SWORD Environment

The students' reviewing activities were enacted using SWORD, a web-based system that students and instructors access from any internet-connected computer using standard Internet browsers (for details, see Cho & Schunn, in press, 2005). SWORD manages the submitting and accessing of drafts, assigning papers to reviewers, submitting and accessing reviews, comparing comments from various reviewers, and passing along feedback to reviewers about the consistency and helpfulness of their comments.

Students logged in to SWORD to access all materials related to the writing and reviewing tasks. After using a word processor of their choice to compose and revise their papers, students submitted the drafts in one of the more accessible formats (i.e., MS-Word, PDF, Rich Text Format, HTML, or simple text). Reviewers were provided access to a randomly selected set of six drafts, generated by a SWORD sampling algorithm. (Instructors are allowed to choose how many drafts are assigned to each peer.) Reviewer ratings and comments for the each of the three ratings dimensions were entered into an online form in SWORD. However, reviewers were encouraged to compose their comments off-line in a word processor and then paste these comments into the online forms. Reviews were made available to writers once the reviewing deadline had passed. When authors had revised their papers and submitted their revised drafts to SWORD, they then had the opportunity to rate the helpfulness of each reviewer's comments within the system.

Comment Analysis

Each reviewer typically generated a number of different comments on each paper on each of the three dimensions (prose flow, argument, and insight). To code comment types, reviewers' comments were first divided into feedback segments (idea units), defined as a self-contained message on a single problem (Artemeva & Logie, 2002). Then each feedback segment was assigned to one of six comment categories: *directive*, *nondirective*, *praise*, *criticism*, *summary*, and *off task*. Table 1 presents the comment-coding scheme with examples.

Table 1
Comment-Coding Scheme, With Definitions and Examples

Type	Definition	Example
Directive	Suggests a specific change particular to the writer's paper.	"In clearly understanding the arguments, I had some trouble. To improve this problem, try to introduce the topics you are going to discuss in your introduction for the reader to connect each point and understand the significance of each argument."
Nondirective	Suggests a nonspecific change that would apply to any paper. Comments on a detail without suggesting a change.	"There were several grammatical and spelling errors that at times made it difficult for me to follow the paper."
Praise	Describes the paper or a portion of the paper positively, including encouraging remarks	"The writing was very catchy and once I started reading the introduction, I just wanted to read even more. Therefore, the introduction is good because it sets the tone of the whole paper making it very insightful and attention getting."
Criticism	Gives a critical or negative evaluation of the paper or a portion of the paper; points out an underdeveloped area. No suggestions for improvement are offered.	"The intro leaves me wondering what the paper is going to be about. At first, the grammatical errors overwhelm the paper, but after putting the pieces together, it is hard to tell if the paper is talking about visual perception, evolution."
Summary	Recapitulates the main points of the paper or a portion of the paper.	"The main points of your paper involve visual perception, the retina, cognitive system, top-down and bottom-up processing, visual data, reading text, and how visual perception affects our everyday lives."
Off task	Comments do not fit any of the code categories; the comments are ambiguous, or a rating was given without written comment.	"I don't know if it was the program or what but the font size was ridiculous so it ended up being short."

Coders were blind to the expertise level of the reviewer. As a check on the reliability of the coding, a random 10% of peer comments were independently coded by a second coder. Agreement between the two coders was

computed as the number of matched segmentations (division of full comments into separate comment segments) and categorizations divided by the number of total segments. Thus, an agreement was recorded only when both segment and assigned category matched for the two coders. The coders reached an acceptable level of 94% agreement.

Design

In one of the three settings, UNDGRD-SM, both the expert and undergraduates produced comments on the students' writing. In this course, we conducted a controlled quasi-experiment in which students received comments from an expert, a single peer, or multiple peers. Consequently, more comments were generated for each paper than were provided to the writers. However, all comments were generated with the expectation that they would be read. The effect of this experimental manipulation on the writers' subsequent revisions will not be discussed in this article (see Cho & Schunn, in press).

In this design, the comments from UNDGRD-SM can be used to compare the feedback from subject-matter expert instructors to feedback from undergraduate peers. The comments produced by undergraduates in UNDGRD-SM and UNDGRD-LG can be used to assess the generality of undergraduate feedback across different course settings. Finally, the comments produced by graduate students can be compared to the comments produced by undergraduates to assess the generality of feedback in different kinds of peer relationships.

It is important to note that all data were generated in the context of psychology classes and may not generalize to other disciplines. However, as a social science, psychology likely sits in the middle of the spectrum of disciplines with respect to how much emphasis is placed on writing, with humanities (e.g., English and history) at one end and natural sciences (physics and chemistry) at the other end.

Results and Discussion

As an overview of writing and reviewing performance, we begin with some simple descriptive statistics about the length of essays and comments that were observed.

Length of Essays

Table 2 presents the average length of the first drafts produced in each class; the variability across classes mainly reflects differences in the assignments

Table 2
Mean Length and Standard Deviation
of First Drafts of Student Essays

Course	Condition	Words		Paragraphs		Pages	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Graduate	Overall	2,510	712	47.6	29.1	9.9	2.3
Large undergraduate	Overall	1,383	357	16.2	7.8	5.5	1.4
Small undergraduate	Single expert	954	123	8.7	1.8	4.1	0.1
	Single peer	1,103	509	17.0	11.7	4.8	1.8
	Multiple peers	840	239	8.3	3.7	3.7	0.9
	Overall	940	368	11.1	8.3	4.1	1.4

to each group, with students in UNDGRD-SM asked only to produce the introductory section of a research report. It is worth noting that both groups of undergraduates wrote essays comparable in length to a typical assignment in 1st-year composition (about 1,000 to 1,400 words).

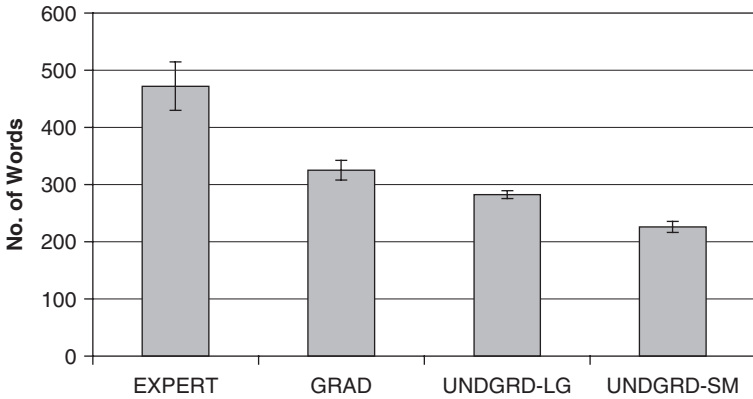
In general, the revised drafts were slightly longer than the first drafts. However, because the most substantive comments were given on the first drafts, only comments on these will be considered further.

Overall Length of Comments

Figure 1 shows the mean length of a reviewer's comments in words. Comments decreased in length at each level of reviewer experience; a one-way analysis of variance (ANOVA) on the number of words with experience status as a between-subject variable showed a significant overall difference, $F(3, 812) = 22.8, p < .001$. The comments from the subject-area expert were longer than those of the graduate students and the graduate students' comments tended to be longer than those of the undergraduates. Bonferroni pairwise analyses show that all the comparisons are significant, $p < .005$, except those between GRAD and UNDGRD-LG.

Figure 2 shows the average number of each type of feedback unit given by a reviewer per writer. A one-way ANOVA revealed that the groups did differ significantly overall in terms of number of idea units, $F(3, 162) = 34.39, p < .001$. More specifically, the expert produced almost twice as many idea units per writer as the graduate and undergraduate students, whereas the student groups produced similar numbers of ideas units—Bonferroni pairwise comparisons show the statistically significant differences were between the expert and each of the others, $p < .005$.

Figure 1
Mean Number of Words in Each Reviewer's Comments (With Standard Error Bars), Prior to Segmenting Comments



Note: 28 reviews were written by the subject-matter expert, 126 reviews by the graduate students, 522 by undergraduates in UNDGRD-LG, and 140 by undergraduates in UNDGRD-SM.

Figure 2
Mean Number of Idea Units in Feedback Per Review by Each Reviewer Type (With Standard Error Bars)

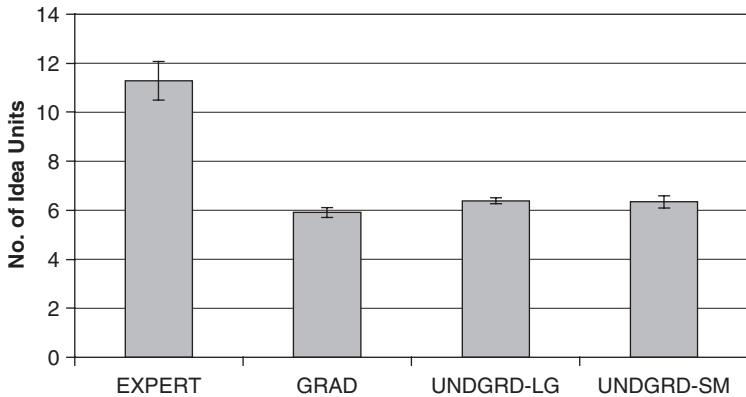
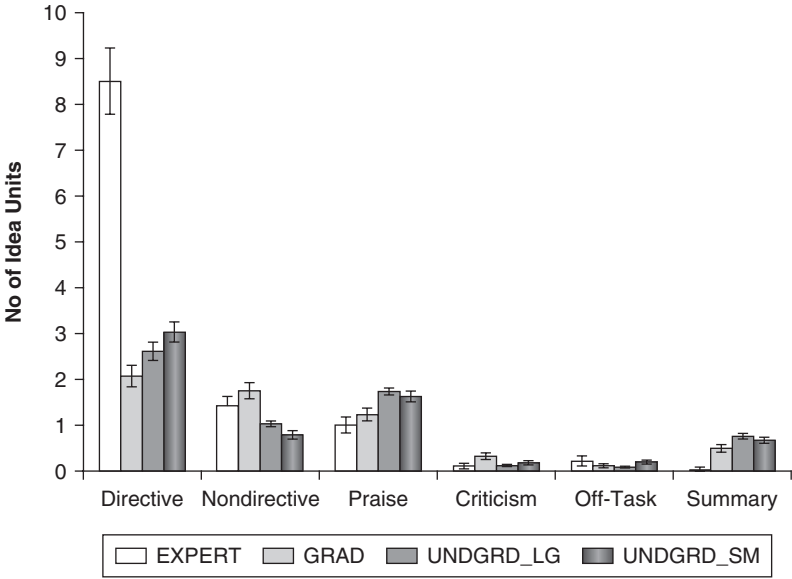


Figure 3
Total Numbers of Idea Units Per Writer in Each Type of Feedback for Each Type of Reviewer (With Standard Error Bars)



The ANOVA on the number of idea units by group was applied to each feedback type (see Figure 3). Statistically significant effects were found for all feedback types except off task: directive feedback, $F(3, 162) = 57.91, p < .001$; nondirective feedback, $F(3, 162) = 10.53, p < .001$; praise, $F(3, 162) = 9.06, p < .001$; criticism, $F(3, 162) = 4.35, p < .01$; summary, $F(3, 162) = 17.42, p < .001$. Bonferroni pairwise comparisons revealed the following differences among groups:

- The expert produced the most directive suggestions, as expected, $p < .005$. This effect is quite large (approximately a 3:1 ratio).
- The expert and graduate students produced more nondirective remarks than the undergraduate students, $p < .005$, perhaps reflecting a better ability to notice more abstract patterns in writing.

- Undergraduates produced more praise comments than the expert (almost 70% more praise than the expert) and graduate students, $p < .005$. Conversely, graduate students produced the most criticism remarks.
- The expert produced the fewest summary statements, presumably reflecting the workload balanced against a belief that summary statements are not as helpful to students.

Amount of Feedback in Each Dimension

In this section, we examine how the overall pattern of comment types within each of the evaluation dimensions (see Figure 4). Note that we do not describe further analyses of off-task and summary feedback because they are infrequent and not of focal interest.

Directive Feedback

Directive feedback involves explicit suggestions of specific changes. For example, one expert comment was

The 3rd paragraph is a collection of 3 seemingly unrelated sources. However, they are all good references and could provide valuable contribution to your arguments. So try to present them in your favor, and build logical connection among these 3 references and between them and your story.

With respect to directive feedback, the expert produced the most comments for all three dimensions of prose flow, $F(3, 162) = 40.35, p < .001$; argument, $F(3, 162) = 42.53, p < .001$; and insight, $F(3, 162) = 10.50, p < .001$, although the difference was largest for the conceptual dimension of argument.

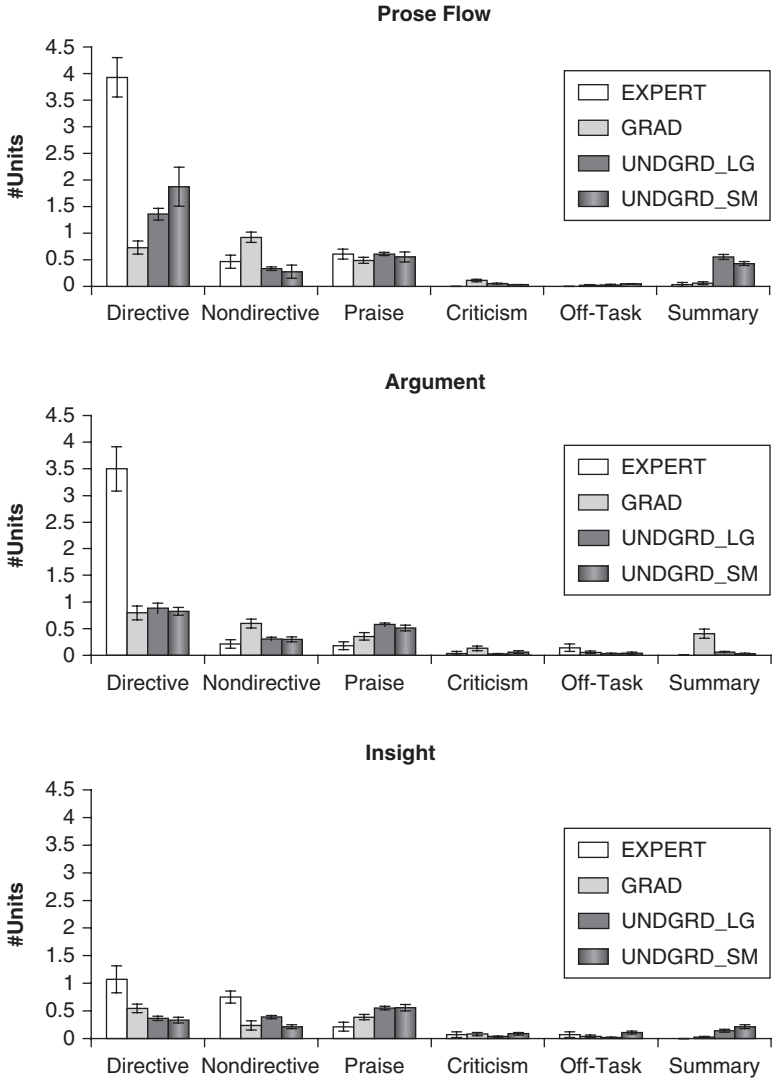
Undergraduates produced more directive comments for prose flow than for argument and insight, $t(165) = 4.30, p < .001$, suggesting that undergraduates focus their detailed comments on the most mechanistic dimension, prose flow.

By contrast, graduate students produced relatively few directive suggestions across the dimensions, suggesting that graduate students either felt that their graduate student peers did not need detailed help in the form of directive feedback or did not have the ability to make detailed feedback for writers at that level.

Nondirective Comments

Nondirective comments make a nonspecific observation that could apply to any paper. For example, an instructor wrote, “Overall very hard to follow

Figure 4
Within Each Evaluative Dimension, the Mean Numbers of Feedback Units (With Standard Error Bars) As a Function of Feedback Type for Each Reviewer Group



the flow, many sentences need to be read over and re-done to be made more clear. I had a very hard time following.”

With respect to nondirective feedback, graduate students had the highest number of idea units for prose flow, $F(3, 162) = 15.32, p < .001$; and argument, $F(3, 162) = 6.98, p < .001$; whereas the expert had the highest number for insight, $F(3, 162) = 13.66, p < .001$. Thus, graduate students put more emphasis on nondirective feedback, especially for prose flow and argument. That the expert had the highest number of nondirective idea units for the insight dimension makes sense considering the very open-ended and domain-knowledge relevant nature of this category.

Praise Comments

Praise comments are defined as encouraging observations on the whole or a portion of the paper. For example, an undergraduate student wrote,

In the closing paragraph you mention the study being conducted near a large research university and hospital system so many of the patrons to local business are college students, employees, and even patients so it makes them more frequently to mention topics related to medicine and other sciences. This is an excellent point and a good way of using previous information to branch off and inquire about other issues relating to your topic.

With respect to praise, undergraduates provided more of these comments on two of the three conceptual dimensions. Undergraduates generated significantly more praise feedback for argument, $F(3, 162) = 12.93, p < .001$, and insight, $F(3, 162) = 9.93, p < .001$. Interestingly, there were no group differences on praise remarks for prose flow—the three levels of reviewers made equivalent numbers of praise comment. These results suggest that undergraduates have a clear comment-giving script that includes giving praise feedback, regardless of the dimensions, whereas experts and graduate students mainly offer praise feedback in the more mechanistic prose flow dimension.

Criticism Comments

Criticism comments express a negative evaluation of the paper or a portion of the paper without improvement suggestions. For example, a student wrote, “The grammar in this paper is extremely poor. It is very hard to follow what is being said.”

With respect to criticism, graduate students produced the most comments on two of the three conceptual dimensions, prose flow, $F(3, 162) = 4.97$, $p < .001$, and argument, $F(3, 162) = 4.17$, $p < .001$. This result supports the notion that graduate students are keen to apply critical thinking skills. It also is consistent with the finding that reviewers focused equally on prose flow in giving praise, whereas the expert and graduate students provided less praise on argument than undergraduates.

Summary

Summary recapitulates the main points or a portion of the paper. For example, an undergraduate student wrote,

The main point of this paper was that you expect a decreased frequency of scientific conversation involving prediction and explanation in a nonscientific setting (coffeehouse), whereas you expect a greater frequency of prediction and explanation to occur in a scientific setting (science museum).

Interestingly, the expert provided no summary on any dimension. Undergraduates provided somewhat more summary comments for the prose flow dimension, whereas graduate students provided somewhat more summary comments on the argument dimension. This tendency is again consistent with other results, indicating a relatively greater emphasis on flow than argument from undergraduates.

Off-Task Comments

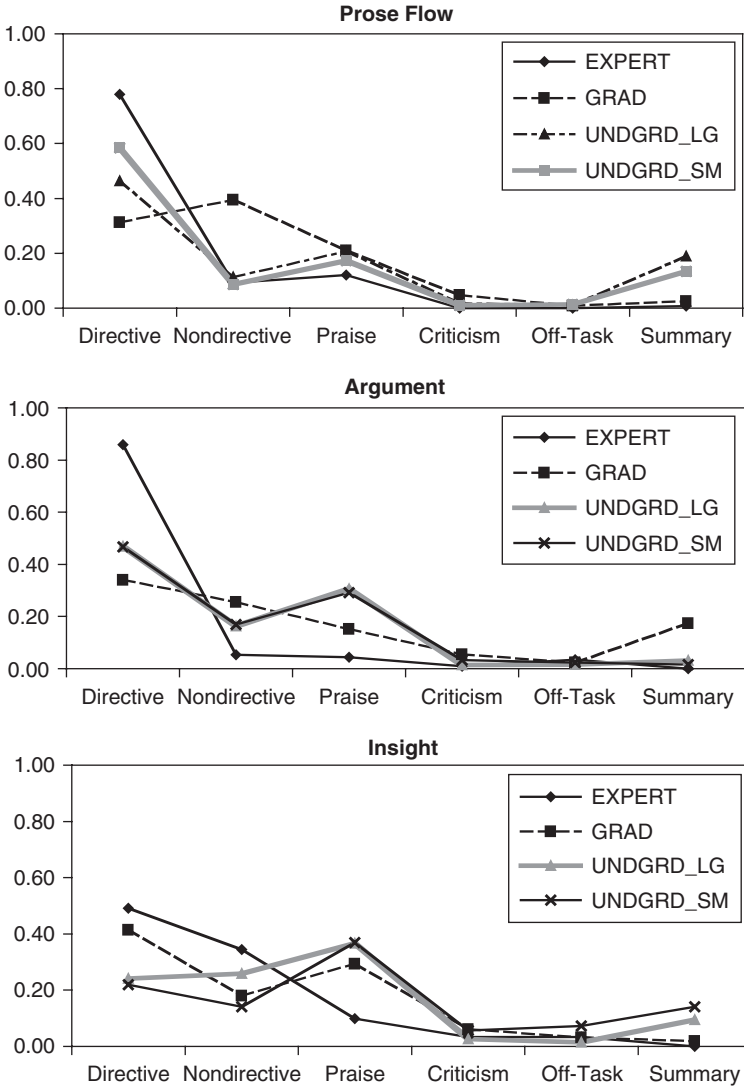
Off-task comments are unrelated to any of the code categories earlier mentioned and unrelated to the student paper. For example, an undergraduate wrote, "I did read your paper, and my original comments will remain as is because I did not see any changes in the text."

Comments of this type were infrequent and equally so in all groups.

Relative Frequency of Feedback

The earlier findings were reexamined in terms of the relative frequency of feedback within each reviewer group (see Figure 5). In other words, instead of examining absolute numbers of each feedback type as a separately analyzed product, we examined the relative emphasis on each feedback

Figure 5
Relative Frequency of Feedback Types in Each
Evaluation Dimension



type within each expertise group. A fairly consistent pattern appeared across the three dimensions of prose flow, argument, and insight.

- Expert feedback was concentrated in directive comments in the prose flow and argument dimensions, and in directive and nondirective feedback in the insight dimension.
- Undergraduate feedback was concentrated in directive and praise comments in all of the three dimensions.
- Graduate student feedback, as might be expected from intermediate experts, sometimes followed the pattern of the expert and sometimes that of the undergraduate students.

Overall, the relative frequency of feedback types followed this order: prose flow—nondirective, directive, then praise; argument—directive then nondirective; and insight—directive then praise.

The relative frequency data provide an opportunity to examine the generality of the undergraduate peer data. As shown in Table 3, the UNDGRD-SM and UNDGRD-LG classes were quite different in several respects. Despite these differences, the correlations between the two patterns were $r = .98$ in prose flow, $r = 1.00$ in argument, and $r = .88$ in insight. Thus, the patterns of peer feedback from the two courses are very consistent in each of the writing evaluation dimensions.

Writer's Perception of Feedback Helpfulness

Expert vs. peer helpfulness. Next we examined whether writers would perceive differences in helpfulness depending upon the feedback source. Using only the small undergraduate class data (instructor and undergraduate feedback on a common set of papers), a two-way mixed ANOVA on the ratings of feedback helpfulness was carried out with all three writing evaluation dimensions as the within-subjects variable and feedback sources as the between-subjects variable. As shown in Figure 6, there were no main effects of expertise source, $F(1, 45) = .86, p = .36$, and evaluation dimension, $F(2, 90) = .97, p = .38$, and no interaction between them, $F(2, 90) = .69, p = .51$. Thus, we found no evidence that undergraduate students thought that peer feedback was significantly less helpful than expert feedback.

Feedback type preference. We also examined what types of feedback student writers consider more helpful than others. To analyze this question,

Table 3
Feature Comparison Between the Experiment
and the Natural Course

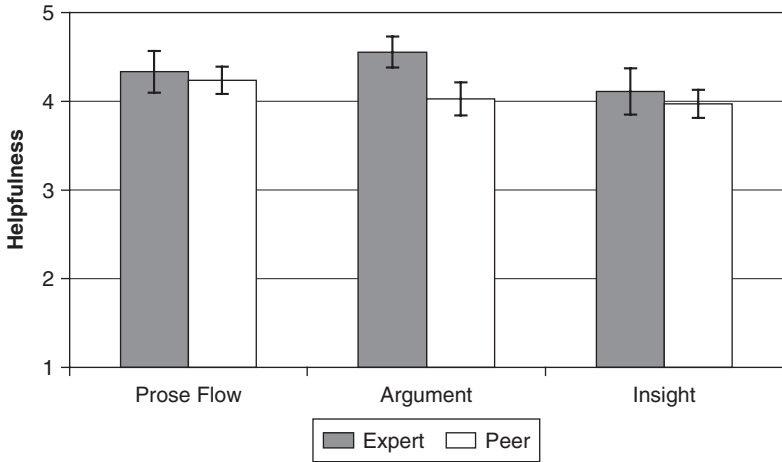
	Small Undergraduate	Large Undergraduate
Course	Research methods	Cognitive psychology for nonmajors
Number of students	28	88
Writing genre	Intro section writing of technical research paper	Academic essay
Manipulation	Experimental control	None = naturalistic

multiple regression analyses were carried out on helpfulness ratings of idea units in each dimension with the six types of feedback as independent variables. For reasons of analysis power, we used data only from the UNDRD-LG course. As shown in Table 4, in all three dimensions, directive and praise feedback positively influenced the perceived helpfulness of the peer review. Critical feedback appeared to have negative influences on perceived helpfulness of comments in argument and insight, whereas summary feedback appeared to have positive influence in prose flow and argument.

General Discussion

The conflict between the need and the practical difficulty of writing practice has led to development of reciprocal peer review systems, such as SWoRD, InSite, and TurnItin.com. Practically, one could argue for the successful contributions to writing instruction of such systems from the better-than-nothing point of view, in that they provide valuable opportunities to practice writing, receive detailed comments, and revise for students in large content courses that may not have a writing component. However, peer review systems may have at least two additional benefits. First, peer review systems provide students with opportunities to critique and reflect upon others' work (Smyth, 2004). Second, peer review systems may produce a more natural writing context; rather than an individual writing to a nonsalient and even unrepresentative reader (i.e., the instructor), a student reads and writes within an active writing community (Applebee, 1981; Cohen & Riel, 1989; Mittan, 1989).

Figure 6
Mean Authors' Helpfulness Evaluations (5 = *Very Helpful*, 1 = *Very Unhelpful*) With Standard Error Bars



Note: The author evaluated how helpful expert or peer feedback was on each evaluation dimension. The figure is based on 9 expert reviews and 38 peer reviews.

Given these developments that are tending to expand the use of peer review systems, the current study provides a useful profile of peer feedback on writing, analyzing the comments of undergraduates relative to those of subject-matter expert, and of graduate students relative to undergraduate students. We summarize our findings in two ways. First, we consider population differences between practically meaningful group contrasts. Second, we consider key dimensions of writing feedback and to what extent our findings contribute to a broader understanding of those dimensions of feedback.

Comments of Subject-Matter Expert Instructors vs. Undergraduate Peers

Most of the subject-matter expert's comments were directive, specifying exactly how to resolve problems. This emphasis on directive feedback held for all three dimensions, from the prose flow dimension, which is

Table 4
Separate Predictiveness (Standardized Coefficients) of Different Comment Types
in a Multiple Regression on Helpfulness Ratings

	Directive	Nondirective	Praise	Criticism	Off-Task	Summary	Model Significance
Prose flow	.24*	-.03	.23*	-.02	-.002	.16*	$F(6, 376) = 8.5, p < .001$
Argument	.15*	.01	.15*	-.20*	-.04	.13*	$F(6, 375) = 6.9, p < .001$
Insight	.11*	.02	.30*	-.14*	-.08	.05	$F(6, 375) = 6.8, p < .001$

* $p < .05$.

knowledge lean, to the insight dimension, which is knowledge rich. By contrast, undergraduate peer reviewers used directive comments much less than the subject-matter expert, especially for the most knowledge-rich dimension, insight. However, undergraduate reviewers did produce large pluralities of directive and praise comments. Undergraduates are especially likely to include praise, which in turn may lead to the improvement of writing quality through motivational effects (Gee, 1972).

Undergraduate Peer Comments in Different Settings

The findings on the undergraduate student comments appear generalizable because the comment patterns from the experimental setting of the small undergraduate class are highly consistent with that from the natural setting of the large undergraduate. Therefore, the profiles of undergraduate peer feedback are likely to be generalizable, at least among psychology students.

Peer Comments by Graduate and Undergraduate Students

Undergraduate peer reviewers' comments feature distinctive use of directive and praise comments, whereas graduate peer reviewers do not use any specific type of comments more than others. The comments from graduate students appear to show transitional pattern between the expert and undergraduate students. In other words, graduate students do not resort to any distinctive types of comments. However, graduate students tend to use criticism noticeably more than undergraduate students who used praise.

Directive vs. Nondirective Comments

As stated earlier, most of the subject-matter expert's feedback was directive, specifying exactly how to resolve problems across all three dimensions. By contrast, student peer reviewers devoted a noticeably smaller proportion of their comments to directive feedback, especially for the most knowledge-rich dimension, insight. In fact, the overall effect of more idea units from instructors was entirely attributable to the much higher number of directive comments, with no other comment types showing higher absolute numbers by the subject-matter expert. However, it should be noted that even peer reviewers did produce a large number of directive comments.

If one assumes that the ideal form of feedback is directive feedback (Beason, 1993; Clare et al., 2000; Coupe, 1986), the type most commonly produced by subject-matter experts, then peer feedback appears moderately valid when little content knowledge is required in reviewing, whereas its validity appears low if rich content knowledge is necessary. Consistent with this expert view of what is ideal, undergraduates consistently rated directive feedback as helpful in all evaluation dimensions, whereas nondirective comments were neutrally evaluated at best. The finding that undergraduates preferred detailed comments with specific suggestions is consistent with Straub (1997).

Criticism vs. Praise Feedback

Consistent with prior work showing the valued nature of praise (Cole, Coats, & Lentell, 1986; Gee, 1972; Seidman, 1968) and mitigating language in feedback (Neuwirth et al., 1994), peers generally rated praise as being very helpful (Straub, 1997). Consistent with these perceptions of helpfulness, undergraduates are especially likely to include praise, which in turn may lead to the improvement of writing quality through motivational effects (Gee, 1972; Seidman, 1968; Stevens, 1973; Taylor & Hoedt, 1966).

By contrast, criticism comments were viewed as very unhelpful (Straub, 1997), especially on the evaluation dimensions of argument and insight. Interestingly, criticism comments were relatively rare, although previous research tended to include negative comments as a major type of feedback (e.g. Gee, 1972; Sproull & Kiesler, 1986). In the currently examined situation, students knew that their feedback comments would be evaluated for helpfulness by the authors. Students were also warned to be constructive rather than punishing in their feedback. These features of the feedback situation may have led to the relative rarity of criticism comments.

Although rare in general, criticism comments were more commonly found in graduate student feedback. Leki (1995) noted that “the more the [individual] knew about the subject matter of the essays, the more critical they were” (p. 32). Another possible factor in producing a relatively high proportion of criticism comments in graduate student feedback is a combination of (a) relatively strong training in critical thinking skills relative to undergraduates and (b) relatively weak practice with giving others feedback relative to a subject-matter expert with considerable prior experience in feedback giving.

Implications for Graduate Student TA-Based or Instructor-Based Writing Instruction

WID courses often take advantage of graduate students as instructors or teaching assistants in reading and giving comments on undergraduate student writing. The finding that comments from graduate students include a higher frequency of criticism has clear implications for graduate student-based writing instruction. Given that graduate students were commenting on papers that were stronger papers than those written by undergraduates and will later be placed in situations of having to provide feedback to undergraduates (e.g., as instructors or TAs), it is likely that graduate students could benefit from some guidance on using more praise and less criticism in their feedback to undergraduates (especially less personal judgment, as recommended by Faigley, 1989).

Implications for WID

To further improve the kinds of feedback that students receive in WID settings, the current findings suggest that instructors should be pushed to include more praise because they appear to rarely provide it, but it appears to be valued by students, even if only from a motivational perspective. By contrast, as instructors begin to rely more heavily on peer comments in WID settings, the current findings suggest that more support needs to be provided for students to be less vague and more directive in their commenting.

Caveats

In this study, reviewers anonymously evaluated manuscripts with the help of the SWoRD system. Anonymity could limit the generalizability of the findings to face-to-face peer reviews. However, anonymity was important for several reasons. According to Freedman (1987), efforts to get students to critique peer writing often fail because peer critiquing violates student social norms of not criticizing other students in the presence of a teacher. In addition, students often respond to authority or status of feedback givers (e.g. Sue-Chan & Latham, 2004). In other words, novices tend to just accept feedback when feedback givers have higher status, whereas they tend to challenge feedback from peers or less-knowledgeable people. Thus, anonymous feedback (as currently studied) may indeed be different from face-to-face feedback, but the anonymous situation may represent the more desirable method for implementing peer feedback.

Another threat to the generalizability of the current study relates to the way subject-matter expert feedback was collected. On the one hand, the subject-matter expert was paid for providing feedback rather than collecting data from the actual course instructor and thus may have provided more feedback to students than a typical course instructor would have. On the other hand, the subject-matter expert was asked to provide feedback on all 28 papers, whereas the students had to provide feedback on only 6 papers. It is important to note that this asymmetry in amount of papers to grade is actually realistic: peers are rarely asked to provide feedback on as many papers as instructors are required to comment upon. At the same time, the pay was instituted to make sure the subject-matter expert was representative of a well-motivated instructor providing real attention to the feedback task.

Furthermore, field-based observations suggest that many instructors struggle in various ways in producing good feedback to students. For example, instructor feedback on student writing often tends to point out problems (e.g., the famous “awk” comment) rather than provide writers with directive feedback on how to fix the problems (Coupe, 1986; Sommers, 1980). In addition, some instructors often use the same feedback on every paper (Sommers, 1980) or focus on mechanical errors such as punctuation (Beason, 1993; Clare et al., 2000). As a result, weak improvement in writing quality has often been attributed to expert or instructors’ nondirective or ambiguous feedback (Beason, 1993; Clare et al., 2000; Coupe, 1986; Sommers, 1980). Thus, it is worth noting that the currently studied subject-matter expert is likely to be typical of good WID instructors, rather than a clear representative of the full class of subject-matter experts.

Future Research

Finally, we note that the current focus has been on the nature of feedback that peers generate. But there are many aspects that should be addressed in future research. In particular, it is important to look more deeply at how students revise their drafts in response to different sources of feedback: peer-only feedback, subject-matter expert feedback, and feedback from writing instructors. Another interesting question is how a student’s reviewing experiences shape his or her own writing later. We are optimistic that the availability of online support systems, such as SWoRD, will facilitate this kind of research.

Appendix A

Writing Prompts for Three Courses

I. Graduate course writing prompt (2002):

The paper can take one of three different forms: (a) a well-argued or supported critique of a central claim made in one of the papers, (b) a discussion of an important new outcome in the topic area, or (c) an important practical application of one of the ideas in the topic area. Do not provide a simple summary of an extra reading—you must show why the critique, outcome, or application is very important. Topic area options:

- Reasoning and rationality
- Architectures of cognition
- Analogy and transfer
- Strategy choice and strategy discovery

II. Large undergraduate course writing prompt (2003)

The goal of the paper is to take some key finding or theory from one of the chapters and apply it to everyday life: Explain something about everyday life and make an argument about how life could be improved using that piece of cognitive psychology, either in terms of how artifacts around us are built, or how education is done, or some other application. Your paper should be five to eight pages long (double-spaced). Your paper will be graded by four to five of your fellow students. You will submit two drafts of the paper, and both drafts will be graded by your peers. Your identify will not be known by the reviewers.

Topic areas (three of eight options):

- *Decision making and deductive reasoning.* How do people make decisions? Do they do it in a rational fashion? What shortcuts do they take in decision making? How do people reason about logical problems? Do they sometimes just make slips or are humans fundamentally illogical beings?
- *Attention.* Attention is about how we focus our efforts on particular aspects of what we are doing, how that changes how well we do a task, and whether there are limits on what we can attend to, how many things we can attend to at the same time, etc. Why do some tasks require lots of attention to do well but others require no effort as well? Why is it possible to do some complicated things such as driving without paying much attention, but sometimes simple tasks require lots of attention?
- *Memory encoding.* To retrieve things from memory at a later date, the information first has to get into long-term memory. Encoding is the

process that puts information into long-term memory. Does everything that gets into primary (or working memory) also make it into long-term memory? Why do people quickly encode lots of information in some settings (such as watching a football game) but are only able to encode a small number of things slowly in another setting (such as mathematics class)?

III. Small undergraduate class writing prompt (Spring 2003):

This task requires you to perform research on the given question. Both in your lab and in this online system, you will develop your research including a written introductory section, experimental design, data collection, data analysis, and final write-up. In this online system, you will submit an introductory section on the topic, write peer reviews of other students' papers, revise your paper based on feedback. In your lab, you will do the other parts of the projects such as research design, data collection, data analysis, and final writing-up.

- *Topic: Informal science learning.* How does context influence the frequency and type of everyday scientific conversation? By context, we mean locations such as a coffeehouse, science museum, etc. By type, we mean description (about data or evidence without linking to theory), explanation (involving theory usually linked to data), and prediction.
- *What to write in your introductory section.* In your intro section, you should state what question your paper is about, why your question is important, what your position or stance is in terms of the literature review, your sound logic and inference, and your hypotheses, what answer you expect for the question.
- *Literature review.* The literature review should give the reader the sense that you have examined some of the material on your topic and are familiar with contrasting perspectives and viewpoints. In this sense, it builds your authority to speak about the issues you are dealing with. It also gives the reader an overview of the issues and problems and what people from various backgrounds generally think about them. In this section, you quote, summarize, and paraphrase what other people say about your topic. Save your own opinions and arguments for the discussion section. In the literature review section, you should describe the sources and the information you found in your research. Organize the material by topic, by perspective, or in some other coherent way. Don't just dump the information in randomly in a "this one says this, that one says that" fashion. By grouping and organizing your information, identifying different issues and positions, and establishing connections and differences between your sources, you provide the reader with an overview of the issue or topic you are writing about.

Appendix B

Instructions for Reviewers

Overview

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

Prose Flow Dimension

Questions: Did the writing flow smoothly so you could follow the main argument? This dimension is not about low level writing problems, such as typos and simple grammar problems, unless those problems are so bad that it makes it hard to follow the argument. Instead, this dimension is about whether you easily understood what each of the arguments was and the ordering of the points made sense to you. Can you find the main points? Are the transitions from one point to the next harsh, or do they transition naturally?

Comments: First summarize what you perceived as the main points being made so that the writer can see whether the readers can follow the paper's arguments. Then, make specific comments about what problems you had in understanding the arguments and following the flow across arguments. Be sure to give specific advice for how to fix the problems.

Argument Dimension

Questions: This dimension is about the logic of the argument being made. Did the author just make some claims, or did the author provide some supporting arguments or evidence for those claims? Did the supporting arguments logically support the claims being made or 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 ignored?

Comments: Provide specific comments about the logic of the author's argument. If points were just made without support, describe which ones they were. If the support provided doesn't make logical sense, explain what they are. If some obvious

counter-argument was not considered, explain what that counter-argument is. Then give potential fixes to these problems if you can think of any. This might involve suggesting that the author change their argument.

Insight Dimension

Questions: This dimension concerns the extent to which new knowledge is introduced by a writer. Did the author just summarize what everybody in the class would already know from coming to class and doing the assigned readings, or did the author tell you something new?

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

Appendix C

A Partial Screen Shot of Review Instructions

1. Prose Flow

Did the writing flow smoothly so you could follow the main argument? This dimension is not about low level writing problems, like typos and similar grammar problems, unless those problems are so bad that it makes it hard to follow the argument. Instead this dimension is about whether you easily understood what each of the argument was and the ordering of the points made sense to you. Can you find the main points? Are the transitions from one point to the next harsh, or do they transition naturally?

Your Comments: First summarize what you perceived as the main points being made so that the writer can see whether the reader can follow the paper's arguments. Then make specific comments about what problems you had in understanding the arguments and following the flow across arguments. Be sure to give specific advice for how to fix the problem.

2. Argument

This dimension is about the logic of the argument being made. Did the author just make some claims, or did the author provide some supporting arguments or evidence for those claims? Did the supporting arguments logically support the claims being made or 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 ignored?

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

3. Insight Beyond Core Readings

This dimension concerns the extent to which new knowledge is introduced by a writer. Did the author just summarize what everybody in the class would already know from coning to class and doing the assigned readings, or did the author tell you something new?

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

References

- Applebee, A. N. (1981). *A study of writing in the secondary school*. Urbana, IL: National Council of Teachers of English.
- Artemeva, N., & Logie, S. (2002). Introducing engineering students to intellectual teamwork: The teaching and practice of peer feedback in the professional communication classroom. *Language and Learning across the Disciplines*, 6, 62-85.
- Atwater, L. A., Waldman, D., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisor's cynicism, follow-up and commitment to subordinates. *Personnel Psychology*, 53, 275-297.
- Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the Teaching of English*, 27, 395-421.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4), 413-427.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474-481.
- Cho, K., & Schunn, C. D. (2003). Validity and reliability of peer assessments with a missing data estimation technique. *Proceedings of ED-Media 2003*, 1511-1514.
- Cho, K., & Schunn, C. D. (in press). Scaffolded writing and rewriting in the discipline. *Computers & Education*.
- Cho, K., & Schunn, C. (2005). The SWoRD system (Version 3.0) [Computer software]. Pittsburg, PA: Learning Research and Development, University of Pittsburgh. Available at <http://ladybug.lrdc.pitt.edu/sword>.
- Clare, L. C., Valdes, R., & Patthey-Chavez, G. G. (2000). Learning to write in urban elementary and middle schools: An investigation of teachers' written feedback on student compositions. *Center for the Study of Evaluation Technical Report*, 526.
- Cohen, M., & Riel, M. (1989). The effect of distant audience on students' writing. *American Educational Research Journal*, 26(2), 143-159.
- Cole, S., Coats, M., & Lentell, H. (1986). Towards good teaching by correspondence. *Open Learning*, 1(1), 16-22.
- Coupe, N. (1986). Evaluating teachers' responses to children's writing. (pp. 122-144). In J. Harris & J. Wilkinson (Eds.), *Reading children's writing: A linguistic view*. London: Allen and Unwin.
- Dragga, S. (Ed.). (1992). *The technical writing repertoire*. Lubbock, TX: Association of Teachers of Technical Writing.
- Faigley, L. (1989). Judging writing, judging selves. *College Composition and Communication*, 40, 395-412.
- Fife, J. M., & O'Neill, P. (2001). Moving beyond the written comment: Narrowing the gap between response practice and research. *College Composition and Communication*, 53(2), 300-321.
- Freedman, S. W. (1987). *Peer response groups in two ninth-grade classrooms* (Tech. Rep. No. 12). Berkeley: University of California.
- Graner, M. H. (1985). Revision techniques: Peer editing and the revision workshop. *Dissertation Abstracts International*, 47, 109.
- Gee, T. (1972). Students' responses to teachers' comments. *Research in the Teaching of English*, 6, 212-221.

- Haswell, R. H. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication, 22*, 198-223.
- Hughes, I. E., & Large, B. J. (1993). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education, 18*, 379-385.
- Johnstone, K. M., Ashbaugh, H., & Warfield, T. D. (2002). Effects of repeated practice and contextual-writing experience on college students' writing skills. *Journal of Educational Psychology, 94*(2), 305-315.
- Leki, I. (1995). Good writing: I know it when I see it. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 23-46). Norwood, NJ: Ablex.
- Magin, D., & Helmore, P. (2001). Peer and teacher assessment of oral presentation skills: How reliable are they? *Studies in Higher Education, 26*, 287-298.
- Mittan, R. (1989). The peer response process: Harnessing students' communicative power. In D. Johnson & D. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 207-219). New York: Longman.
- Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing—A case-study from geography. *Innovations in Education and Training International, 32*, 324-335.
- Neuwirth, C. M., Chandook, R., Charney, D., Wojahn, P., & Kim, L. (1994). Distributed collaborative writing: A comparison of spoken and written modalities for reviewing and revising documents. *Proceedings of the Computer-Human Interaction '94 Conference, April 24-28, 1994, Boston Massachusetts* (pp. 51-57). New York: Association for Computing Machinery.
- Richer, D. L. (1992). The effects of two feedback systems on first year college students' writing proficiency. *Dissertation Abstracts International, 53*, 2722.
- Rucker, M. L., & Thompson, S. (2003). Assessing student learning outcomes: An investigation of the relationship among feedback measures. *College Student Journal, 37*, 400-404.
- Seidman, E. (1968). Marking students' compositions: Implications of achievement motivation theory. *Dissertation Abstracts International, 28*, 2605-A.
- Shrauger, J. S., & Schoenemann, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin, 86*, 3, 549-573.
- Smith, S. (2003). The role of technical expertise in engineering and writing teachers' evaluations of students' writing. *Written Communication, 20*, 37-80.
- Smith, S. (1997). The genre of the end comment: Conventions in teacher responses to student writing. *College Composition and Communication, 48*, 249-268.
- Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment & Evaluation in Higher Education, 29*(3), 370-378.
- Sommers, N. I. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*, 378-387.
- Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organization communication. *Management Science, 32*, 1492-1512.
- Stevens, A. E. (1973). The effects of positive and negative evaluation on the written composition of low performing high school students. *Dissertation Abstracts International, 34*, 1778-A.
- Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English, 31*, 91-119.
- Sue-Chan, C., & Latham, G. P. (2004). The relative effectiveness of external, peer, and self-coaches. *Applied Psychology: An international Review, 53*(2), 260-278.

Taylor, W., & Hoedt, K. (1966). The effect of praise upon the quality and quantity of creative writing. *Journal of Educational Research*, 60, 80-83.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.

Kwangsung Cho received his PhD in cognitive psychology from the University of Pittsburgh in 2004. He is an assistant professor in the School of Information Science and Learning Technologies, at the University of Missouri, Columbia. His research focuses on understanding the nature of peer collaboration in writing instruction and developing computer-supported writing programs.

Christian D. Schunn received his PhD in psychology from Carnegie Mellon University in 1995. He is currently a research scientist at the Learning Research and Development Center and assistant professor of psychology, cognitive studies in education, and intelligent systems at the University of Pittsburgh. His current research focuses on understanding complex forms of expertise, building models of authentic practice in science and engineering, and applying those models to improve K-20 science education.

David Charney is a professor in the Division of Rhetoric and Writing at the University of Texas at Austin. Her review of research on audience, persuasion, and argument with Carolyn Miller will appear in the *Handbook of Writing Research*, edited by Charles Bazerman. Her argumentation textbook, coauthored with Christine Neuwirth, David Kaufer, and Cheryl Geisler, *Having Your Say*, was published this year by Pearson Longman.