

Students' perceptions about peer assessment for writing: their origin and impact on revision work

Julia H. Kaufman · Christian D. Schunn

Received: 18 February 2008 / Accepted: 11 March 2010
© Springer Science+Business Media B.V. 2010

Abstract We investigate students' negative perceptions about an online peer assessment system for undergraduate writing across the disciplines. Specifically, we consider the nature of students' resistance to peer assessment; what factors influence that resistance; and how students' perceptions impact their revision work. We do this work by first examining findings from an end-of-course survey administered to 250 students in ten courses across six universities using an online peer assessment system called SWoRD for their writing assignments. Those findings indicate that students have the most positive perceptions of SWoRD in those courses where an instructor graded their work in addition to peers (as opposed to peer-only grading). We then move to an in-depth examination of perceptions and revision work among 84 students using SWoRD and no instructor grading for assessment of writing in one university class. Findings from that study indicate that students sometimes regard peer assessment as unfair and often believe that peers are unqualified to review and assess students' work. Furthermore, students' perceptions about the fairness of peer assessment drop significantly following students' experience in doing peer assessment. Students' fairness perceptions—and drops in those perceptions—are most significantly associated with their perceptions about the extent to which peers' feedback is useful and positive. However, students' perceptions appear to be unrelated to the extent of their revision work. This research fills a considerable gap in the literature regarding the origin of students' negative perceptions about peer assessment, as well as how perceptions influence performance.

Keywords Peer assessment · Peer feedback · Peer review · Higher education · Student perceptions · Student attitudes · Resistance · Technology · Writing · Revision

J. H. Kaufman (✉) · C. D. Schunn
Learning Research and Development Center, University of Pittsburgh,
3939 O'Hara Street, Pittsburgh, PA 15260, USA
e-mail: juliak@pitt.edu

National reports have documented students' poor performance on national writing assessments and the dearth of writing instruction across all education levels (Graham and Perin 2007; National Commission on Writing in American Schools and Colleges 2003; Persky et al. 2003). These reports particularly cite the lack of teacher time to evaluate writing and little use of technologies that might enable more quality writing instruction across disciplines. Online peer assessment is one technological tool that could considerably cut teachers' time in assigning and grading writing, and such assessment has been proven as reliable and valid (Cho et al. 2006; Rushton et al. 1993). However, studies have also highlighted students' discomfort with peer assessment and their reservations about its fairness and accuracy (Cheng and Warren 1997; Davies 2000; Liu and Carless 2006; Rushton et al. 1993; Smith et al. 2002). Students' resistance and negative perceptions could discourage teachers from using such technological innovations (Cohen 1988; Jacobsen 1997; McNeil 1988).

While research demonstrates students' resistance to peer assessment, it provides only tentative and anecdotal evidence about the factors that encourage that resistance (Davies 2000; Liu and Carless 2006; Smith et al. 2002). Additionally, while some studies demonstrate that students' perceptions influence their approach to study (Crawford et al. 1998; Lizzio et al. 2002; Ramsden 1991), little work specifically addresses how perceptions impact revisions of writing (Simkin and Ramarapu 1997). This study therefore delves into students' perceptions about online peer assessment and particularly students' perceptions that such assessment is unfair. In our analysis, we consider both the factors that influence those perceptions and how students' perceptions relate to their revision work. We do this work by first examining general findings from an end-of-course survey completed by 250 students in 10 courses across six universities using an online peer assessment system called SWoRD for their writing assignments. We then move to an in-depth examination of perceptions and revision work among 84 students using the same online peer assessment system for writing and revision work in one university class.

Peer assessment and student resistance

Much research has demonstrated the positive benefits of peer assessment for both the assessor and student who is receiving the assessment. Falchikov (1986) and Roscoe and Chi (2007) note that students who assess the work of their peers are engaging in a cognitively-demanding activity that extends their understanding of subject matter and writing. For the student who receives peer review, studies report deepened subject matter knowledge (Barak and Rafaeli 2004; Venables and Summit 2003) and a more positive attitude about writing (Katstra et al. 1987).

Research documents students' concerns about the fairness of peer assessment (Cheng and Warren 1997; Rushton et al. 1993) and the lack of instructor input in the peer assessment process (Liu and Carless 2006; Sluijsmans et al. 2001; Smith et al. 2002). In their study, Smith et al. (2002) specifically note that although students communicated a higher level of confidence in the peer assessment process over time and continual experience with peer assessment, "the unease about fairness and consistency [of peer assessment] remained" (p. 76).

In contrast to studies regarding students' negative opinions about peer assessment, some research studies have documented students' general satisfaction and positive perceptions of the peer assessment process. Gatfield (1999), for example, noted students' general satisfaction with a peer-assessed marketing project, while Wen and Tsai (2006) documented

positive perceptions about peer assessment among university students in Taiwan. However, neither case specifically measured students' perceptions of fairness about peer assessment. Furthermore, Wen and Tsai also noted that students had a less favorable reaction to the idea of online peer assessment compared to other peer assessment methods, and they expressed little confidence in the validity of online assessment.

Even if students were provided with objective evidence about the validity of peer assessment, that evidence may not persuade students to view peer assessment in a more positive light. In his review of the peer assessment research, Topping (1998) noted that students' acceptance of peer assessment is quite independent from their knowledge about the demonstrated reliability and validity of that assessment. Cho et al.'s (2006) work to demonstrate the reliability and validity of an online peer assessment system—SwoRD—offers further validation of Topping's observation. In their study, they examined data from 708 students across 16 different courses (large and small, undergraduate and graduate) from 4 universities and found that the aggregate ratings of at least four peers on a piece of writing are as reliable and valid as a single instructor rating. However, in five of these courses, students gave low ratings of reliability and validity for the peer assessment process.

While research documents students' attitudes toward peer assessment, such research offers no empirical data on what factors impact those attitudes, beyond some qualitative observations. For example, Davies (2000) and Liu and Carless (2006) note that some students who have negative perceptions of peer assessment doubt the expertise of their fellow students (as compared to their instructors). Researchers have also hypothesized that students' discomfort and negativity can be traced to the problematic power relations that students associate with assessing their peers (Falchikov 2001; Liu and Carless 2006; Smith et al. 2002). These studies indicate that students may have trouble with the non-traditional idea of their peers assessing their writing in place of an instructor. Furthermore, variation in students' attitudes about assessment may rely a great deal on how individual instructors introduce and plan peer assessment. However, research does not provide conclusive evidence regarding the origin of students' attitudes toward peer assessment.

Studies have also advocated for certain steps to alleviate students' negative perceptions of peer assessment, including more peer assessment experience (Sluijsmans et al. 2001; Wen and Tsai 2006); clarity about peer assessment criteria (Falchikov 2005; Smith et al. 2002); and support and training in regard to the peer assessment process (Cheng and Warren 1997; Falchikov 2005, 2007). Berg et al. (2006) have further suggested that specific elements of the peer assessment context might lead to better revisions and a more positive attitude toward peer assessment, including small assessment groups of 3–4 students and face-to-face contact among assessors and those who are being assessed. However, these prescriptions are not based on clear evidence about the roots of students' anxiety and negativity about peer assessment. Thus, the research highlights a need for more inquiry into students' perceptions about peer assessment and what drives those perceptions.

Peer assessment, students' perceptions and their performance

Beyond the need to know why students have negative perceptions of peer assessment in order to design better peer assessment processes, we also need more evidence about how peer assessment, and students' perceptions of peer assessment, impact the quality of their work, particularly in the realm of higher education where student satisfaction becomes increasingly important for course evaluations and the classes students choose to take. While some higher education research indicates that students' satisfaction with school and

positive perceptions of their learning environment influences their GPA (Bean and Bradley 1986; Lizzio et al. 2002), studies specifically focused on peer assessment have demonstrated that students' discomfort with peer review does not correlate with their grades (Simkin and Ramarapu 1997).

Furthermore, despite evidence of students' generally negative opinions of peer assessment, research consistently indicates that feedback from multiple peers can yield improvements in writing that are significantly higher than improvements for those students who received assessment only from an instructor or subject-matter expert. For example, in comparing writing improvements among two college classes—one class who received only peer feedback prior to revisions and one class who received only instructor feedback prior to revisions—Richer (1992) found significantly more writing improvement among those in the peer feedback only group. Additionally, in their research on writing improvements among university students in a Research Methods class, Cho and Schunn (2007) noted significantly more writing improvement among a group of students who received feedback on their writing drafts only from six peers compared to another group of students who received feedback on their drafts only from a subject-matter expert.

In these cases where students see improvements to their writing following peer review, the improvements could be the result of a *higher quantity* of feedback from multiple peers as opposed to instructor feedback from only one source. In his review of the literature on peer assessment, Topping (1998) allows that “peer feedback might not be of the high quality expected from a professional staff member, [however] its greater immediacy, frequency, and volume compensate for this” (p. 255). Thus, the benefits of peer assessment could center upon the ability of multiple peers to produce an overall evaluation that is comparable to or better than that of a single instructor.

While improvement in students' writing could be due to the higher quantity of feedback that students receive from multiple peers compared to one instructor, improvements might also be linked the more positive nature of peer feedback. While Zhao (1998) found that peers tend to be more critical in the context of anonymous peer review, both Cho et al. (2006) and Patchan et al. (2009) demonstrated that peers using SWORD—which is an anonymous peer assessment system—gave about twice as many praise comments to peers compared to instructors. Cho et al. (2006) further noted that students found comments containing praise and mitigating language to be more helpful than other types of comments. At the same time, research on SWORD has not been able to link praise and mitigating language to actual improvements in students' writing (Nelson and Schunn 2008).

Apart from studies on peer assessment, students' perceptions of peer assessment and their writing revisions, some higher education research has considered how students' perceptions drive surface (oriented toward reproduction for tests) versus deep (oriented toward understanding) approaches to studying. These studies note that students with positive perceptions of teaching and course goals often adopt a deep approach to studying, whereas students with negative perceptions about course workload and the appropriateness of assessment often adopt a surface approach to studying (Crawford et al. 1998; Lizzio et al. 2002; Ramsden 1991). However, this research is based solely on students' self-reports and not empirical measures of students' work.

In contrast to the suggestions that positive perceptions drive revision behavior and deep versus surface learning approaches, Bures et al. (2000) found that students with negative impressions about the usefulness of computer conferencing would still participate in the conferencing if it was an important part of their course grade. Their work suggests that perceptions may matter less for performance if students are concerned about getting a high

grade. However, research has not demonstrated a tie between students' revision work and their concern about grades, nor have studies specifically examined how peer evaluation connects with students' revision work.

In light of the gaps in the literature in regard to what factors drive students' negative perceptions about peer assessment and how those perceptions impact students' revision behavior, we investigate undergraduate students' opinions about an online system for peer review and assessment called SWoRD (Scaffolded Writing and Reviewing in the Discipline) and their revision behavior in response to peer assessment. We specifically focus on the nature of students' perceptions about SWoRD and what variables influence those perceptions, and we examine how students' perceptions impact their revision work. To do this work, we briefly consider students' end-of-course survey responses regarding SWoRD across six universities, and we consider the factors that might influence students' responses. Then, we consider data from an intensive study in a single class, examining both the factors that may influence students' perceptions of SWoRD and how their perceptions affect their revision work.

As per the literature on peer assessment, we conjecture that some students will have negative attitudes in regard to peer assessment and the fairness of peer grading, and we posit that students' negative or positive attitudes about peer assessment will be independent of any information they receive about the demonstrated validity and reliability of SWoRD. Additionally, we posit that those negative attitudes are connected with students' opposition to peers assessing their work in place of an instructor. Furthermore, following the literature on students' perceptions and their work in school, we hypothesize that students' negative attitudes about peer assessment may impact their revision work. Such a relationship between students' perceptions and revisions would imply that much more work must be done to help students see peer assessment in a more positive light. This paper thus examines both what drives students' attitudes toward peer assessment and how those attitudes impact students' work in order to inform better design and implementation of peer assessment and other student-centered technologies.

Method

Participants

For our initial examination of end-of-course surveys, participants were 250 undergraduate and graduate students enrolled in 10 courses in various disciplines across 6 universities, each taught by a different instructor sometime in 2007 or 2008. The students used SWoRD to submit draft and final versions of at least one paper and no more than three papers. The classes varied in size (10–220 students, mean = 65) and they also varied in regard to (1) whether the instructor took peer assessment into account when calculating students' grades; and (2) whether the instructor also graded papers alongside peers. For all these classes, students completed an end-of-course survey on their perceptions of SWoRD and peer assessment (at return rates ranging from 15 to 73%). Because survey length often influences response rates, additional demographics on survey participants were not collected.

For our more intensive study of students in one course, the participants were undergraduate students with varied majors enrolled in a Fall 2006 cognitive science course at a large public university in the United States. The students used the online SWoRD process to submit the draft and final versions of two papers, as well as to review and assess the work of their peers for those two papers. The writing prompt for each paper asked students

Table 1 Demographic information for participants in intensive one-course study

	Survey participants (<i>n</i> = 84)	Interview participants (<i>n</i> = 15)
Sex		
Female	55	9
Male	29	6
Year in school		
Freshman	33	10
Sophomore	21	1
Junior	16	2
Senior	12	2
Unknown	2	0

to describe a cognitive science theory and its applicability to everyday life, taking peers in their class into account as the intended audience for their writing. See Table 1 below for participants' demographic information; the table includes 84 students who participated in both a pre- and post-survey in regard to their perceptions and use of SWoRD, as well as the 15 students who (in addition to completing a survey) completed one semi-structured interview about their experiences with SWoRD.

Materials

SWoRD (Scaffolded Writing and Reviewing in the Discipline)

SWoRD is a Web-based client–server application that allows students to submit their first drafts online; each draft is then distributed to five students who review and evaluate their peers' work. SWoRD defines reviewing as a constructive process of determining how well the text represents the writer's intentions and knowledge, and the reviewers engage in this process through the evaluation of typically three default dimensions: flow (understandability of main points and transitions from one point to another), logic (organized and coherent structure of facts and arguments, including how well main arguments are supported by evidence), and insight (original contribution to new knowledge and ideas). The reviews that students submit back to their peers must include written feedback, as well as a rating of the draft's quality on a seven-point rating scale (1: Disastrous to 7: Excellent) for each dimension.

After the reviewers have submitted their comments and numerical ratings, SWoRD uses a formula taking into account systematic differences, consistency, and spread of scores to determine the accuracy of each student's review score; the system then weights each score accordingly. When all reviews for a paper have been received, SWoRD provides students with their peers' feedback and a weighted rating of their draft.¹ After receiving their reviews, students rewrite their papers and turn in a final draft, which is reviewed and evaluated again by the same peers who assessed students' first draft. After SWoRD weights the second set of scores, the final reviews and ratings are provided to the authors. For our study, students engaged in this writing-reviewing-revising-reviewing process twice, once for each paper that they wrote.

¹ Both the author and reviewers are assigned pseudonyms during this process. Thus, the reviewers do not know who wrote a given paper, and authors do not know who reviewed their paper.

Additionally, after authors have received their reviews and allowed some time for revision and/or reflection, they also provide their reviewers with feedback about the helpfulness of the reviews, again using a seven-point rating scale with written comments. This allows reviewers to receive constructive feedback about the usefulness of their reviews. The SWoRD system is thus designed so that students can work to improve themselves both as writers and critical assessors of each others' work.

End-of-course survey for initial study

For our initial, brief analysis, we asked instructors to administer an end-of-course survey to students across 10 courses using SWoRD. The survey asked for students' level of agreement with a series of statements about SWoRD: "SWoRD improved the quality of my writing"; "SWoRD improved the quality of my reviewing"; "I got a fair grade through peer evaluation"; "My writing improved through revisions I made in response to peer comments"; "I think it is reasonable to be given grades for peer reviewing activities"; "I think it is reasonable to use peer ratings for assigning grades to papers"; "I think this course should keep using SWoRD"; and "I think other courses should start using SWoRD." Because factor analysis indicated that responses to all eight statements varied together (Eigenvalue = 6.01), students' responses were combined into one composite overall rating of SWoRD ($\alpha = .92$).

Semi-structured Interviews for Intensive One-Course Study

Our interviews with students using SWoRD in one course, which lasted anywhere from 20 to 60 min, included a variety of set questions and probes that helped students describe their experiences in using the SWoRD system and elaborate on their writing process. To understand students' experiences with SWoRD, we asked some questions that were similar to survey items, including questions about their qualifications and their peers' qualifications to give feedback and grades. Questions about students' writing process addressed their past experiences with peer assessment, their preparation for writing papers, the content of their papers, and their revision process.

Pre- and post-surveys for intensive one-course study

The pre- and post-surveys asked for students' agreement with statements about the usefulness of peers' and one's own feedback; the reliability and validity² of peers' and one's own feedback; the positive nature of peers' and one's own feedback; and the fairness of peer grades. Table 2 lists the specific survey questions for constructs used in our data analysis of students' perceptions, and—when a composite was used to measure fairness—the alphas for that composite. Beyond measuring students' perceptions about peer assessment, the surveys also contained questions about students' prior experience with peer review, SAT score,³ writing ability,⁴ sex, year in school, and their perceptions about the

² Because peers in the class were the intended audience for students' writing, survey measurements of validity focused on how well students' feedback aligned with other peers' feedback.

³ For students who reported ACT scores, SAT is predicted through a concordance formula using ACT scores (see Dorans 1999).

⁴ In the survey, students' writing ability was assessed in two ways: students were asked questions both about their perceptions of their writing ability (on a scale from 1 (poor) to 5 (excellent) and the writing class

Table 2 Survey items measuring constructs used in data analysis

Construct	Statement (students asked to respond whether they strongly agree, agree, disagree, strongly disagree, or are unsure)
Usefulness of own feedback	The feedback I (give/gave) my peers on their writing for this class (will be/was) useful
Positive nature of own feedback	The feedback I (give/gave) my peers on their writing (will likely be/was) too negative or critical. [Agreement reverse coded for this item]
Validity of own feedback	The feedback I (give/gave) a peer on his/her paper probably (will be/was) similar to the feedback that other peers (give/gave) on the same paper
Reliability of own feedback	If I had to give feedback several months from now on the same papers for which I (will give/gave) feedback in this class, I would probably give similar feedback
Usefulness of peers' feedback	The feedback my peers (give/gave) me on my writing for this class (will be/was) useful
Positive nature of peers' feedback	The feedback peers (give/gave) me on my writing (will likely be/was) too negative or critical. [Agreement reverse coded for this item]
Validity of peers' feedback	The feedback I (get/got) from one peer (will be/was) similar to the feedback I (get/got) from other peers on the same paper
Reliability of peers' feedback	If my peers gave me feedback several months from now on the same paper they (will examine/examined) for this class, they would probably give me similar feedback
Fairness of peers' feedback (composite of two statements)	Peers (will give/gave) me a fair grade on my writing I (will receive/received) a fair average grade on my writing through the grades given to me by multiple peers ($\alpha = .76$ for these two items in pre-survey and $\alpha = .78$ in post-survey)

Parentheses indicates wording in pre-survey versus post-survey item

importance of grades. The post-survey also incorporated a few questions not addressed in the pre-survey, including some questions about peers' and one's own qualification to give feedback and grades and an open-ended request for students' "thoughts about SWoRD and/or about peer review in general."

We triangulated our data in multiple ways to establish convergent validation of our survey instruments. First, we ran correlations between students' level of agreement for post-survey statements in regard to the qualifications of students and their peers (e.g. "My peers are qualified to give me a grade on my writing"; 1 = Strongly disagree and 5 = Strongly agree) and the qualitative coding for students' response to the same statements asked as questions in the interview (e.g. "Are your peers qualified to give you a grade on your writing?" coded as 1 = No, 2 = Unsure or Maybe, 3 = Yes). We found positive, moderate correlations between post-survey and interview responses ($n = 14$), which are as follows: Peers are qualified to comment on my writing/Are your peers qualified to give you feedback on your writing? (.32); Peers are qualified to give me a grade on my writing/Are peers qualified to give you a grade on your writing? (.33); I am

Footnote 4 continued

in which they had placed at the university (when students are freshmen, the English Department places them in a remedial, regular, or advanced writing class, depending on a variety of factors that include grades in high school, SAT scores, and a writing sample that they submit as a freshmen at the university).

qualified to provide comments to peers on their writing/Are you qualified to give feedback on your peers' writing? (.46); I am qualified to comment on my peers writing/Are you qualified to comment on your peers' writing? (.26).

While the positive correlations between survey and interview responses were encouraging, the sample size was too small to draw any strong conclusions about the validity of the survey. We therefore also looked at the correlation between the students' attitude toward SWoRD as expressed in their post-survey open-ended comment (coded as 1 = Negative; 2 = Neutral; 3 = Positive) and their overall perceptions of SWoRD as expressed in their level of agreement to questions about the usefulness, validity, reliability, and fairness of their comments and their peers' comments. We found this correlation to be very high (.55, $p < .01$), which also suggests that the survey questions are at least providing an accurate gauge of students' attitude toward SWoRD.

Peer assessment scores and revision work

In order to link survey and interview responses with students' performance, we took into account students' raw evaluation scores from their peers on draft and final papers, as well as the final grades calculated through the SWoRD system. Additionally, we considered students' revision work as an indication of their effort to improve their writing after receiving their peers' comments. Revision work is represented as the simple and complex changes that students make to their work, which were calculated through a software program that counts changes made from one paper draft to the next (including insertions, deletions, substitutions, etc.). If changes involved four or more consecutive words, they were coded as complex changes. Changes of three or fewer consecutive words were coded as simple changes. We investigated the validity of our measure by examining the relationship between revision work and gains in students' peer assessment score from draft one to final draft. For each paper, the count of complex changes is significantly correlated ($p < .01$) with the gains in students' scores (.45 for paper one and .39 for paper two). The correlations are present but not significant between the count of simple changes and gains in peer assessment score from draft one to draft two (.23 for paper one and .07 for paper two). Thus, the complex changes appear to be likely a good indicator of quality revision work.

Additionally, we found that students' number of simple changes for their revision of paper one was very significantly correlated with their number of simple changes for their revision of paper two (.45, $p < .01$), as was students' number of complex changes for their revision of paper one and their revision of paper two (.44, $p < .01$). At the same time, simple changes in the first paper are not significantly related to complex changes in the second paper; nor our simple changes in the second paper related to complex changes in the first. These finding suggests that students' propensity to make "simple changes" or "complex changes" is stable over time and thus that our measures for simple and complex changes are reliable.

Procedures

For our initial study, end-of-course surveys were administered to students in ten undergraduate and graduate courses after they had completed their experience using SWoRD for peer assessment of their writing. We only included responses in our dataset when we had received surveys from at least ten students from within a class.

For our intensive study of one course, surveys and interviews were administered to students during the Fall 2006 semester at the same time that students used the SWoRD process to write drafts and final versions of two papers. The pre-survey was administered before students began their first paper, and the post-survey was administered after students received a final grade from their peers on their second paper. Interviews took place a little over midway through the semester, after students had received both their draft and final feedback and assessment from peers for their first paper. At the end of the semester, students' scores and revision work were calculated separately, and those data were matched with student survey and interview data.

In our coding and analysis, we moved back and forth among our data sources in order to triangulate our findings (Denzin 1978; Tashakkori and Teddlie 1998). For example, after noting students' negative feelings about peer assessment in student interviews, we examined their alignment with students' closed-ended and open-ended responses in the survey regarding the fairness and usefulness of peer assessment. The qualitative interview data were transcribed and coded according to major categories of questions, as well as according to emergent categories suggested by initial reading and analysis of the interviews (Strauss and Corbin 1998). For instance, after preliminary interview data coding, we noticed a trend in regard to some students who were more reflective and thoughtful about the peer assessment and revision process, and we thus did additional coding and analysis to understand how that reflectivity presented itself across student interviews. Coding of those qualitative data directed us to pay attention to certain variables in the survey analysis, particularly the reservations or uncertainty expressed by almost all students in regard to the fairness of peer assessment.

The survey analysis explored changes in students' perceptions from the pre- to the post-survey (using paired *t* tests); correlations among pre-survey items, post-survey items, and changes from pre- to post-survey; and multiple linear regressions focusing on what factors specifically influences students perceptions of fairness in the pre- and post-survey, as well as declines in perceptions of fairness from the pre- to post-survey. We conducted additional quantitative analysis to consider whether perceptions measured in the survey—as well as other demographic characteristics—were correlated with students' revision work. Significant differences and correlations in the quantitative analysis directed us back again to particular areas of the qualitative data in order to uncover additional emerging categories for analysis. Thus, for the analysis, the qualitative and quantitative data work together to address issues of general causation and provide rich, detailed explanations and examples for the questions under investigation.

Results and discussion

For our initial study, students were asked in an end-of-course survey to express their agreement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree) with statements about how SWoRD has influenced their writing and whether the SWoRD process is fair, reasonable and accurate. Of these statements, students' expressed the lowest agreement with the statements, "I think it is reasonable to be given grades for peer reviewing" (2.3 on a scale of 1–5) and "I think it is reasonable to have peer feedback instead of instructor feedback" (2.6). They expressed the highest agreement with statements, "My writing improved through revisions made in response to peer comments" (3.5) and "I got accurate comments from peers" (3.5).

We ran by-course correlations ($n = 10$) between students' overall perceptions of SWoRD and other factors related to the specific context in which SWoRD was used by students. Those factors included whether students received a grade for peers' assessment of their writing (true in 6 of 10 courses), whether students' final grades also included instructor assessment of students' written work (true in 3 of 10 courses), number of students in the course (range of 10–220 students, mean = 65), and the number of papers in a course (two or more papers in 3 courses; one paper in 7 courses). The highest correlation was between students' positive perceptions of SWoRD and whether the instructor also evaluated their written work (Pearson correlation = .63, p -value = .05). The number of required papers in the course was also highly negatively correlated with students' positive perceptions of SWoRD (Pearson correlation = $-.51$). The number of papers in a course may be mildly conflated with whether an instructor evaluated students' work, as three of the seven courses in which the instructor did not assess students' writing in addition to peers were also courses where students wrote two or more papers (in the seven remaining courses, the instructor asked students to write only one paper).

We should also note that students' positive perceptions of SWoRD were positively (although non-significantly) correlated with courses in which peer feedback counted toward students' grades at all (Pearson correlation = .33). Thus, peer grading itself does not seem to be an issue leading to students' negative perceptions of SWoRD. Instead, grading by both peers and the instructor appears to lead to more positive perceptions, as opposed to grading by peers only.

Given that students had the most negative perceptions of SWoRD in courses where only peers (and not the instructor) evaluated students' work, we may understand more about the nature of students' negative perceptions by focusing on a course where the instructor did not assess students' written work and only relied on peer assessment to issue grades for student writing. We therefore now move to an intensive examination of students' perceptions and revision work in such a course.

To explore the findings from that intensive study, we first look at students' perceptions about online peer assessment. Then, having located perceptions of fairness as an issue causing the bulk of students' negative attitudes toward SWoRD, we hone in on what factors are influencing those perceptions of fairness. Finally, we explore how students' perceptions, as measured in the survey, impact their revision work.

Context for peer assessment in one-course study

In this course, grades for writing and revision of the two assigned papers came only from undergraduate students who used SWoRD to provide writing ratings and feedback to one another. As with all peer review through SWoRD, students provided anonymous feedback to their peers and did not know the identities of the writers for whom they provided feedback and ratings. In the class, the instructor took about 30 minutes of class time to provide students with a brief verbal description of SWoRD, describe its reliability and validity, and discuss common tendencies for giving feedback that students should avoid (e.g. being too critical or uncritical with peer feedback and assessment). Additionally, students took a brief online tutorial on how to give good peer feedback (Available at: <http://www.pitt.edu/~schunn/feedback/>). The instructor generally introduced SWoRD as a valuable tool that would provide students with useful feedback to help them improve their writing.

Students' reservations about peer assessment

As in prior studies noting students' negative perceptions of peer assessment, many students in this study expressed some negative feelings about the fairness of peer assessment using SWORD. While most of the students with whom we spoke in interviews conceded that five peers are as likely to give an accurate and reliable grade as one instructor, every student but one nonetheless expressed some reservation about themselves or their peers being qualified to assess student writing. Most of those reservations had to do with the idea of peers providing feedback and, particularly, grades to one another without a professor's input. For example, in response to a question about whether peers are qualified to give feedback, a student stated, "No, because I'm not qualified and if they're even, if the class even consists of 50% of people like me, I don't think they're qualified at all." Similarly, in response to a question about whether he feels qualified to give peers a grade, another student commented, "I wouldn't say that I feel qualified [to grade my peers]... maybe just because I'm used to getting grades from teachers... I just don't think that I personally have enough experience in reviewing and critiquing papers."

These negative responses about peer assessment are echoed in students' open-ended response to a post-survey question about the SWORD process. In that open-ended response, almost 30% of the responding students ($n = 24$) made comments about SWORD that can be categorized as mostly negative, while another 32% ($n = 27$) expressed at least some uncertainty or reservations about using SWORD that were not outweighed by their positive comments. For instance, although one student allowed that SWORD "was an alright system," he continued onto say, "I still am unsure whether it is really fair to be graded by other kids in the class." The negative tenor of these comments is echoed by students in other classes and at other universities who have used SWORD for peer assessment. While these data do not demonstrate that the majority of students who use SWORD dislike the SWORD process, findings do suggest that the majority of students who use SWORD will have doubts about its usefulness, which is of particular concern for teachers who wish to use SWORD or other peer assessment processes in their classrooms.

Beyond evidence that some students are concerned about the fairness of SWORD, more concerning are the significant drops in students' perceptions of fairness that are reflected in pre- and post-survey comparisons. In Fig. 1 below, differences in pre- and post-survey responses demonstrate these decreases in students' agreement about getting fair grades from peers, along with decreases in students' agreement that peers' feedback is useful and valid. As can also be seen in the table, drops in regard to fairness, usefulness of peers' feedback, and validity of peers' feedback occur even despite a slight rise in perceptions about the reliability and positivity of peers' feedback and students' perception of their own feedback as reliable, useful, and valid. This finding echoes Smith et al.'s (2002) observation that students may be reasonably confident about the reliability of the peer assessment process after engaging in peer assessment and yet still feel some unease about giving and getting grades from their peers. In this case, the data indicate that students' increased unease after using peer assessment for two papers centers around doubts about the fairness of peer assessment, as well as uncertainty about the usefulness and validity of their peers' feedback.

Factors impacting perceptions about fairness of peer assessment

Because most students express at least some reservations uncertainty about the fairness of peer assessment, and because students' perceptions about the fairness of peer assessment

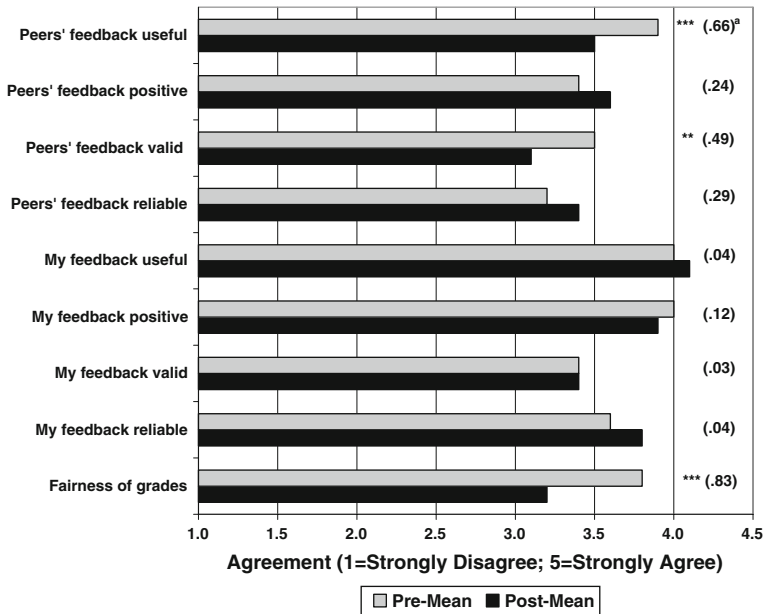


Fig. 1 Paired *t* test of students' pre- and post-survey responses. ** *p* < .01; *** *p* < .001. ^aEffect size [effect size = mean difference/standard deviation of pre-survey mean]

Table 3 Significant pearson correlations with pre-survey perceptions of fairness and pre-survey to post-survey drops in perceptions of fairness

Pre- and post-survey factors	Pre-fairness	Post-fairness		Drops in fairness
Peers will give (gave) useful feedback	.14	.59**	Drops in perceptions of peers' useful feedback	.41**
Peers will give (gave) positive feedback	.42**	.36**	Drops in perceptions of peers' positive feedback	.39**
Peers will give (gave) valid feedback	.21	.25*	Drops in perceptions of peers' valid feedback	.26*

* *p* < .05; ** *p* < .01

dropped very significantly from the pre- to the post-survey, this section explores the factors related to students' fairness perceptions, as calculated through correlations and regressions. Table 3 explores correlations of other survey items with fairness and drops in fairness. As discussed previously, the fairness variable takes into account students' agreement that they received fair grades from their peers. Only three survey items were significantly correlated with fairness: usefulness of peers' feedback, positive nature of peers' feedback, and validity of peers' feedback. However, those three items were significantly correlated with perceptions of fairness across time points. In Table 3, the left column presents pre-correlations (pre-survey measurements of fairness with other pre-survey items); the middle column presents post-correlations (post-survey measurements of fairness with other post-survey items). The right column presented drop correlations (drops in fairness with drops in other items from pre-survey to post-survey).

Table 4 Unstandardized betas and (standard errors) for multiple linear regressions on fairness

	Pre-survey fairness perceptions B (SE)	Post-survey fairness perceptions B (SE)	Drops in fairness perceptions B (SE)
Peers will give (gave) useful feedback	.10 (.11)	.45*** (.08)	NA
Peers will give (gave) positive feedback	.34** (.10)	.19 (.09)	NA
Peers will give (gave) valid feedback	.15 (.10)	.02 (.08)	NA
Drops in perceptions of peers' useful feedback	NA	NA	.25* (.11)
Drops in perceptions of peers' positive feedback	NA	NA	.24* (.10)
Drops in perceptions of peers' valid feedback	NA	NA	.08 (.11)
Overall R ²	.21	.38	.24

* $p < .05$; ** $p < .01$; *** $p < .0001$

As can be seen in the table, pre-survey perceptions about fairness are most closely tied with pre-survey agreement that peers will give positive feedback. However, in the post-survey, agreement with items about the usefulness and validity of peers' feedback—in addition to agreement that their peers will give them positive feedback—correlate most highly with perceptions of fairness. Similarly, drops in all three factors—perceptions about the usefulness, positivity, and validity of peers' feedback—are all correlated with drops in perceptions of fairness. Thus, while perceptions about the usefulness of peers' feedback do not play a role in perceptions of fairness before students have done peer assessment, they appear to play the largest role in perceptions of fairness *after* students have engaged in peer assessment.

The non-significant correlations with fairness for other survey items are not noted in the correlation table, and those non-significant items include past experience with peer review, SAT scores, self-assessments of writing ability, prior writing classes and grades in those classes, sex, year in school, grades that students received through peer assessment, and students' perceptions about their own feedback and the reliability of their peers' feedback. Students' negative opinions about the fairness of peer assessment thus seem to be primarily tied with students' perceptions about the quality of their peers' feedback.

In order to best understand how perceptions about peers' feedback work together to impact perceptions of fairness, we also performed multiple linear regressions on fairness perceptions. Table 4 presents the results for multiple regressions on (1) students' pre-survey perceptions about the fairness of grades through peer assessment (using pre-survey items as independent variables); (2) students' post-survey perceptions about the fairness of grades through peer assessment (using post-survey items as independent variables); and (3) drops in fairness perceptions from the pre- to post-survey (using drops in items from pre- to post-survey as independent variables). As can be seen in the table, students' pre-survey perceptions about the fairness of grades through peer assessment are most largely driven by students' agreement that their peers will give them positive feedback. However, in the post-survey, after students have completed all their peer assessment tasks in the course, students' fairness perceptions are strongly driven by whether they agree that their peers

have given them useful feedback. Similarly, drops in fairness perceptions are most connected with drops in perceptions of peers' useful feedback alongside drops in perceptions that peers gave students positive feedback.

These findings suggest that the content of the feedback that students receive (or their perceptions about the content of that feedback) may be the most important factor in determining whether students think that peer assessment is fair after they have gone through the peer assessment process. Thus, while students might hold more traditional ideas about the need for an instructor to assess their work (versus their peers), their concerns about the fairness of the grades they get from peers (and drops in those perceptions of fairness) have much to do with whether they think that they got positive, useful feedback from their peers.

Perceptions and revision behavior

In this final section, we explore whether students' perceptions about peer assessment impact their revision work. If negative perceptions about the fairness, usefulness, and validity of peer assessment influence students' motivation and effort to improve their writing, then we must question the value of peer assessment more carefully. For our analysis, revision work takes into account the simple changes (three consecutive words or less, not including function words) and complex changes (four consecutive words or more) that students make from draft one to draft two of each paper.

To consider the relationship between revision work and perceptions, we ran correlations for simple and complex changes on each paper alongside all students' perceptions as measured on both the pre-survey and post-survey. To further explore other factors that might influence revision work, we also performed correlations between paper changes and students' demographic factors measured in the survey, as well as students' SAT and writing scores. However, we found no significant or noteworthy correlations between paper changes and any factors we measured in the survey.

Additionally, we considered whether changes in revision behavior from the first to the second paper were associated with changes in perceptions. In this analysis, we particularly focused on associations with students' complex changes, since—as mentioned in our methodology section—those complex changes are highly associated with gains in students' grades from their first to second draft of each paper and are therefore likely an important indicator of excellent revision work. In comparing revisions for the first and second paper using paired *t* tests, we found that students' number of complex revisions decreased significantly from the first to second paper ($p < .01$). Their simple revisions also decreased, although not significantly. However, while—as mentioned above—we saw drops in students' positive perceptions regarding the usefulness, validity, and fairness of their peers' feedback, these drops and any other changes in perceptions were unrelated to changes in students' revision work.

Students' work therefore does not appear to be influenced by students' opinions—negative or positive—in regard to peer assessment. Survey measures for demographic characteristics and students' perceptions about the importance of grades also appear not to influence their revision work, aside from a small but significant correlation between simple changes that students made on their first paper and their post-survey rating of the importance of grades for graduation from college ($r = .25, p < .05$).

These findings from our quantitative data mirror evidence in our qualitative interviews about the lack of a relationship between perceptions and revision work, and an examination of those qualitative findings may shed more light on the factors that do and do not influence

revision behavior. Table 5 below lists the fifteen students who were interviewed about their experience in writing papers using SWoRD. For each student, we list the number of complex changes that each student made to their first paper alongside their opinion (yes, unsure or maybe, no) about whether peers are qualified to give comments and grades, which was derived from coding of their interview responses. Complex changes are listed for the first paper only, since interviews only took place after students had completed revisions for their first paper and had not yet begun revising their second paper.

As with the quantitative data, these qualitative data suggest no particular pattern between students' opinions on the qualifications of their peers and the number of complex changes that they make to their paper. For further explanation on what drives revision work, we looked closely at those students who had made a higher number of complex changes and considered patterns of response in their interviews. We found that the students who made the most changes generally talked in a more reflective way about how the SWoRD process guided improvements and revisions to their paper, and their level of reflection was not necessarily connected to their impression of SWoRD.

Simone, who is a freshmen, is one example of a student who had a fairly negative opinion about SWoRD and yet still made a high number of complex changes on her paper. In fact, Simone made more complex changes on her first paper than any other student in her class. However, she was particularly adamant that peers should not grade one another, stating, "I don't feel that's right at all... they have the right to give me feedback on how to make my paper better, and then the ultimate grade is left up to my professor. But for determining my grade for the class, I don't think they have the qualifications to do that." While this comment indicates that she feels peers have the "right" to give her feedback but not grades, she later added that if other students are "in the same situation I am where this is one of their first big writing classes in college, I don't feel that they have the right [to give feedback] at all." Similarly, she also felt that she was not qualified herself to give

Table 5 Complex revisions on the first paper and perceptions about peers' qualifications to give comments and grades

Pseudonyms for interviewed students	Number of complex paper changes	Are peers qualified to give comments?	Are peers qualified to give grades?
Sam	2	Yes	Yes
Andy	3	Unsure/Maybe	Yes
Brenna	5	Yes	No
Chris	6	Yes	Unsure/Maybe
Bob	6	Yes	Yes
Kevin	7	Yes	Unsure/Maybe
Lisa	11	No	No
Samantha	12	Yes	Yes
Melissa	16	No	No
Wendy	17	Yes	Unsure/Maybe
Dana	21	Yes	Yes
Tara	23	Unsure/Maybe	Yes
Fiona	35	Yes	Yes
Ernie	37	Yes	Unsure/Maybe
Simone	40	Unsure/Maybe	No

feedback or grades: “I don’t think I’m qualified to give [students] feedback that will affect their numeric grade just because my feedback could be something that’s completely wrong.”

That said, her negative perceptions of SWoRD seemingly have nothing to do with her revision work. Instead, her revision work may reflect Simone’s high degree of reflection about what she should improve in her paper and why she should improve it. For example, when speaking about reviewing the work of her peers, Simone noted that the review process “actually helped me figure out... how I wanted to make my paper better... for example, there was one paper that didn’t define a certain word that they needed to define. It was, like, the stronghold of their paper, and I realized, ‘I didn’t do that either. Oh my goodness, I’m probably going to get the same review that I’m giving out.’” Additionally, when explaining how she revised her paper, she considered how a naïve reader—much like the peers who would be reading one another’s work—might view her writing. For example, she said that when she revised she “elaborated more on detail and made sure some of the complex wording was taken out... [complex wording] didn’t really accomplish much other than confusing the reader.” Interestingly, even while Simone professes disagreement with the SWoRD process, her words here demonstrate that she derived some benefits that came specifically from the peer assessment process, including an understanding about where she could improve by reading others’ papers and attending to how readers might perceive her work.

As with Simone, other students who revised a great deal also exemplified a relatively high degree of reflection about wanting to clarify particular points in their writing, as well as discussing how the act of reviewing helped to improve their writing. However, also as with Simone, the act of revision was independent of students’ feelings about the qualifications of peers to give feedback and assessment. Thus, in both the qualitative and quantitative data, whether students liked or approved of peer assessment did not affect the quality of their work within the peer assessment system. This is particularly good news for those who wish to implement peer assessment in their own classrooms, as well as those who are struggling to introduce other innovations in classroom environments where students resist such innovations.

Implications and conclusions

In this paper, we demonstrated that students have particular concern about the fairness of online peer assessment, which is reflected in other studies about peer assessment. However, our work also moves beyond prior studies by delving into the factors that influence perceptions about the fairness of peer assessment and how perceptions impact students’ revision work. In our broad examination of end-of-course surveys among students who have used SWoRD, we found that students’ negative perceptions of SWoRD were strongest in those courses where grades for writing were provided only by peers and not by the instructor. As per the literature, we find that students’ perceptions stand apart from any demonstration of the reliability and validity of peer assessment using SWoRD.

These findings suggest that instructors might allay students’ concerns about the fairness of peer assessment by participating in the grading process. However, SWoRD was specifically designed to provide university students with more quality writing experiences without over-burdening instructors who would otherwise be unable to offer such experiences. We thus worked to further understand students’ resistance to peer-only assessment by embarking on an in-depth study of one course in which grades for writing were

provided only by peers. In that study, we found that students' initial concerns about the fairness of peer assessment (prior to using peer assessment) were most influenced by their concerns about the extent to which peers' feedback will be positive. However, after students engaged in peer assessment for two separate papers, their perceptions about the usefulness of peer assessment became the largest factor connected with both their perceptions of fairness and drops in those fairness perceptions. Thus, in the course that we examined, the peer assessment experience itself—and possibly the content of feedback that students received from their peers—was the most salient issue influencing negative attitudes about the fairness of SWoRD.

More encouraging is the finding that students' negative perceptions appear not to impact their revision work. This finding runs counter to research that students' positive impressions of their learning environment—including the teaching and appropriateness of assessments—seem to drive a deeper approach to study and that negative impressions drive a more surface approach. Additionally, concern about one's grades does not appear to guide revision behavior in any strong way. Qualitative data suggest that students' degree of reflection about their writing and about the benefits of the reviewing process played a role in extent of their revision work. That reflection may be prompted by students' intrinsic motivation to learn or some other factor unmeasured by the present study.

While students' negative perceptions appear not to influence their revision work, those negative perceptions are nonetheless a major theme in our data. The student concerns expressed in this data also arise in student evaluations for instructors who are using SWoRD. Because such student evaluations are often the primary method for university instructor evaluation and may discourage instructors from using much-needed innovations, it is therefore important to consider how SWoRD and other such technological innovations might be better designed to appeal to students.

As with previous work on peer assessment, our work indicates that ongoing support and training (Falchikov 2005, 2007; Smith et al. 2002)—particularly training focused on helping students to give positive, constructive feedback—could improve students' perceptions of peer assessment. As we noted, the primary issue driving students' perceptions of fairness in our one-course study appears to be whether peers' comments were useful and positive, and the peer assessment experience over the course of two papers led to drops in perceptions of fairness. It thus stands to reason that a training that enables students to give more positive, useful feedback to their peers would likely improve students' perceptions of fairness regarding the peer review process. In our study, students did participate in online tutorial on how to give positive and constructive feedback to their peers, including how to be more specific about the location of problems in a paper and how to offer viable suggestions for fixing those problems. However, we do not know the extent to which students found the tutorial convincing or influential for their reviewing practices, and further research could shed more light on whether this or more in-depth support improves students' perceptions about peer assessment.

Other studies on peer assessment suggest that students' continued exposure to peer assessment (Sluijsmans et al. 2001; Wen and Tsai 2006) will help them to view that assessment more positively. At first glance, our research appears to run counter to this assumption because students' perceptions about the fairness of peer assessment declined following their experience with peer assessment. However, these declines only came after students used peer assessment for the writing of two papers in a single course. This study therefore does not necessarily imply that students' perceptions will also decline if they have more peer assessment experiences across a variety of courses. Through the use of peer assessment in more classes, students may begin to regard such assessment as a normal part

of their education and may also understand more clearly how their peers' advice can contribute to their education. Continued use of peer assessment may also encourage students to see themselves as one justifiable audience for their peers' writing and thus a valuable source of feedback about that writing.

In addition, future research should examine the role that other individual differences play in student acceptance of peer feedback, such as motivational factors or resilience to critical feedback. Finally, studies across settings and subjects could provide more information about how students' perceptions influence the work they do and allow for some generalizations about how students' attitudes affect their performance. Specifically, research that examines student perceptions and their performance in relation to other student-centered technologies could pave the way to better design and implementation of such technologies for students at all levels of education.

References

- Barak, M., & Rafaeli, S. (2004). On-line question-posing and peer-assessment as a means for web-based knowledge sharing in learning. *International Journal of Human-Computer Studies*, 61(1), 84–103.
- Bean, J. P., & Bradley, R. K. (1986). Untangling the satisfaction-performance relationship for college students. *Journal of Higher Education*, 57(4), 393–412.
- Bures, E. M., Abrami, P. C., & Amundsen, C. (2000). Student motivation to learn via computer conferencing. *Research in Higher Education*, 41(5), 593–621.
- Cheng, W., & Warren, M. (1997). Having second thoughts: Students' perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233–239.
- Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the disciplines: A web-based reciprocal peer review system. *Computers & Education*, 48(3), 409–426.
- Cho, K., Schunn, C. D., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891–901.
- Cohen, D. K. (1988). Educational technology and school organization. In R. S. Nickerson & P. P. Zohdhiates (Eds.), *Technology in education: Looking toward 2020* (pp. 231–264). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). Qualitatively different experiences of learning mathematics at university. *Learning and Instruction*, 8(5), 455–468.
- Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Teaching International*, 37(4), 346–355.
- Denzin, N. K. (1978). The logic of naturalistic inquiry. In N. K. Denzin (Ed.), *Sociological methods: A sourcebook*. New York: McGraw-Hill.
- Dorans, N. J. (1999). Correspondence between ACT and SAT scores. In *College Board Research Report 99-1*. New York, NY: The College Board.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessment. *Assessment and Evaluation in Higher Education*, 11(2), 146–166.
- Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London: Routledge Falmer.
- Falchikov, N. (2005). *Improving assessment through student involvement*. New York: Routledge Falmer.
- Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 128–143). New York: Routledge.
- Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment and Evaluation in Higher Education*, 24(4), 365–377.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high school—a report to the carnegie corporation of new york*. Washington, DC: Alliance for Excellent Education.
- Jacobsen, D. M. (1997). *Instructional quality, student satisfaction, student success, and student evaluations of faculty: What are the issues in higher education?* Calgary, Alberta: University of Calgary (retrieved January 14, 2008, from ERIC Document Reproduction Service E*Subscribe, ERIC No. ED 423 786).
- Katstra, J., Tollefson, N., & Gilbert, E. (1987). The effects of peer evaluation on attitude toward writing and writing fluency of ninth grade students. *Journal of Educational Research*, 80(3), 168–172.

- Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.
- Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*(1), 27–52.
- McNeil, L. M. (1988). *Contradictions of control: School structure and school knowledge*. New York: Routledge.
- National Commission on Writing in American Schools and Colleges. (2003). *The neglected r: The need for a writing revolution*. New York, NY: College Board. Available from <http://www.writingcommission.org/report.html>.
- Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 1573–1952.
- 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.
- Persky, H. R., Daane, M. C., & Jin, Y. (2003). The nation's report card: Writing 2002 (nces 2003-529). US Department of Education. Institute of Education Sciences. National Center for Education Statistics. Washington DC: Government Printing Office.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education, 16*(2), 129–150.
- Richer, D. L. (1992). *The effects of two feedback systems on first year college students' writing proficiency*. College of Education University of Massachusetts-Lowell.
- Roscoe, R. D., & Chi, M. T. H. (2007). Tutor learning: The role of explaining and responding to questions. *Instructional Science*. Online First.
- Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment. *Journal of Computer-Based Instruction, 20*, 75–80.
- Simkin, M. G., & Ramarapu, N. K. (1997). Student perceptions of the peer review process in student writing projects. *Journal of Technical Writing and Communication, 27*(3), 249–263.
- Sluijsmans, D. M. A., Moerkerke, G., van Merriënboer, J. J. G., & Dochy, F. J. R. C. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation, 27*(2), 153–173.
- Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*(1), 71–81.
- Strauss, A., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage Publications, Inc.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications, Inc.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.
- van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. *Assessment & Evaluation in Higher Education, 31*(1), 19–36.
- Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International, 40*(3), 281–290.
- Wen, M. L., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 27*(18), 27–44.
- Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunications, 4*(4), 311–345.