Peer review in ESL/EFL writing courses: A look at five studies

Peter E. NEFF

Keywords: Peer review, language teaching, second language acquisition, composition

Abstract

In recent years, much attention has been paid to the use of peer review in English writing courses. While the concept of allowing students to evaluate and comment on each other’s written output may intuitively make sense to many language teachers, especially as a component of instruction based on the increasingly influential social-constructivist ideas of Lev Vygotsky (1978, 1986), it is also important that the effectiveness of using such activities be periodically examined. Since the early 1990s a number of studies have been undertaken by second and foreign language English instructors in an effort to determine the results of adding peer review to the essay-composing curriculum. This article reviews several of these studies in an effort to arrive at any overall conclusions which can be drawn about incorporating peer review into the writing process for English language learners. In this critical analysis, the perceived strengths and weaknesses of each study are examined as well as considerations of different types of peer review setups. The paper finishes with several recommendations for future
studies.

**Introduction**

In recent decades there has been a significant shift in English composition classrooms from a focus on “product”—the completed written assignment that students hand in—to “process”—the steps taken by students on the way to creating a final draft. The initial context for this change occurred in L1 writing classrooms, primarily due to the influential ideas first proposed by Peter Elbow (1973) and later endorsed by other researchers (Flower & Hayes, 1981; Hillocks, 1986, Kroll & Vann, 1981). Interest in the journey from conceptualization to finished product has spawned a number of studies, many of which have sought to analyze various stages in the writing process and the effectiveness (or lack thereof) of different approaches which could be utilized by the writing teacher. One method which has often been employed is the use of the students themselves to read and make critical evaluations of one another’s work. This practice is often justified on the grounds that it will help students learn to recognize a potential audience in their peers rather than the teacher alone. Elbow (1973) contended that students become better able to communicate through their writing when considering their peers’ feedback and then revising with them in mind. Many of the theoretical underpinnings for this view derive from, or are compatible with, the writings of Vygotsky (1978, 1986), who considered social interaction to be a crucial element in the development of an individual’s cognitive capacity to learn.

Throughout the 1970s and 1980s most of the research carried out in this area centered on native English-speaking classrooms, with occasional studies being done in English as a Second Language (ESL) contexts. From the early 1990s, however, a greater interest was taken in how much benefit second and foreign language learners could derive from this process of “peer review” (also called “peer response” and “peer evaluation”) in terms
of writing improvement. Researchers with an interest in second language writing began to ask questions about the use of peer review activities that could not always be adequately answered using existing native-speaker studies: Are second language students capable of criticizing each other’s work? If peer suggestions were offered, what kind would be most common? And would the writer then take heed of those comments and revise their writing accordingly? What modes and setups work most effectively? And finally, how can students be effectively prepared to engage in such an activity?

In reality there are no simple answers to any of these questions, but a growing number of second language researchers have attempted to consider their ramifications by developing different strategies for using peer review in second language educational contexts. This critical literature review examines the progress that has been made in our understanding of the role of peer review in the L2 writing classroom since such research first started to be published in major academic journals. Five of the most widely read and cited studies on the peer review process, which examine various factors involved in using such activities in ESL/EFL contexts, and which document the development of peer review research over the last 15 years, have been selected. For each study, an overview of the researchers’ questions, the context, data collection procedures, and conclusions are provided after which follows a critical analysis of the strengths and potential weaknesses of the research, including aspects of the design, participants, analysis, and interpretations. Finally, an attempt is made to tie the findings from the various articles together in order to draw any conclusions about the state of research in this area. Ultimately it is hoped that, through examination of these studies, some consensus can be reached on the place of peer review activities in second language writing curriculums. Additionally, suggestions for future studies in this relatively new line of inquiry, as well as comments on the author’s ongoing research, will conclude this analysis.
Critical Review

Jane Stanley (1992)

Summary

This often-cited study, titled “Coaching student writers to be more effective peer evaluators,” is the earliest of those being examined and is one of the first major examples of peer review research being done within an L2 context—in this case a university ESL writing class. Using previous literature in both first language (L1) and second language (L2) studies as a basis and justification for introducing peer evaluation into two advanced-level ESL composition classes at an American university, Stanley set out to discover what actually occurs during sessions involving “peer-evaluation groups”—small groups of students who read and then orally comment on each other’s writing. From there she attempted to find out whether extensive time spent training (“coaching”) the students beforehand to become effective evaluators would result in better communication strategies used during evaluative sessions. To this end, two different classes were used for the study—one which received extensive training (7 hours in total) on becoming effective evaluators, and the other which received only rudimentary training (about 1 hour’s worth). Both groups were small—fewer than 20 students—and aside from the first group receiving more extensive training in evaluation methods, both classes were taught in much the same way.

During the first group’s training, students were given samples of previous students’ work at different stages of development—from first to final draft—and asked to comment on them as well as pinpoint any parts which seemed vague or contained communicative gaps. Later-stage essays were also looked at for signs of revision. Once this was done, students were asked to consider the strengths or weaknesses of each essay and
contemplate how best to communicate their feedback to the writer. Extensive use of role plays was utilized for this part of the training, with two students pretending to be the writer and the reader. These role plays were performed in front of the class, and observers were asked to comment on which strategies were most effective. From this activity, some consensus came about as to communicative strategies that worked best (specific and short comments were preferred). In contrast to this extensive training, the second group was introduced to the idea of peer evaluation by watching the two writing instructors do a 25-minute role-play of a peer review session which was then followed by a brief class discussion on the topic.

After students were coached as to how to be effective peer responders, they engaged in writing 6 essays over the duration of the course which were later read by their classmates. The students were then asked to comment on each other’s work in discussion pairs. These discussion sessions were recorded and later transcribed and analyzed for categorization based on the type of comments that were made by both evaluator and writer. Turn-taking patterns were analyzed as well as the length of each turn in order to compare the word-length evaluator and writer comments.

In her Findings section, Stanley stated that the group who received extensive training recorded significantly more responses during peer review sessions than did the group who received less training (623 vs. 137). Additionally, she found that the well-trained groups’ comments had a higher incidence of specific response types (“pointing,” “advising,” and “collaborating”) than did the other group. She also determined that the first group’s members had gained more confidence in their roles as evaluators based on their greater use of “announcing” and “reacting” strategies and that writers were also more engaged because of the high incidence of “eliciting” strategies in this group. Additionally, there was more balance between writer and evaluator comments in the coached group, whereas the evaluators in the uncoached group tended to dominate discussions. Stanley
finished detailing her findings by briefly explaining the results of evaluative comments on students’ subsequent drafts. Once again the coached group’s revisions based on comments were greater than those of the uncoached group (68 revisions vs. 16). In her conclusion to the study, Stanley points to the use of extensive coaching as having real merit in improving students’ abilities as peer evaluators.

Analysis

This was a well-organized study which benefited from a solid methodology. Stanley obviously gave consideration to many issues before beginning the study which allowed her to reach more concrete conclusions about the significance of coaching students to be effective evaluators as well as encouraging them to consider communication strategies when put into a peer review scenario. The setup of her study was thoroughly explained, and the fact that she was able to use two groups of similar students in similar circumstances was an obvious benefit. Furthermore, the coding system she used for categorizing student comments—including seven types for her “Evaluator Response Code” and four types for her “Writer Response Code”—seemed appropriate for the research conducted in that they were straightforward and appeared to capture most of the classes of interactions that would be of interest to the researcher. Her analysis of comments using this system was thorough, and having a high degree of inter-rater reliability (92%) was testament to the strength of the system. Additionally, the coaching activities she engaged in with the students seemed beneficial as a way of helping them self-discover the most effective ways to approach their roles as responders and writers. Stanley clearly spent a considerable amount of time preparing for this training, and her results appear to bear out the positive results of engaging in such activities before proceeding with a similar peer review exercise. In sum, there are many strengths in this study in terms of design and implementation, and future
studies into peer review would do well to follow many of Stanley’s coaching techniques.

A few concerns should also be noted, however. First, the limited number of students—only 12 received extensive coaching and 18 were in the uncoached group—undoubtedly affects the veracity of the conclusions which can be drawn from the results of this study. While smaller groups are easier to work with, train, and record, data analysis would have benefited from using a larger sample size. With writing from only 12 students being analyzed, each participant’s writing tendencies would seem to have a significant effect on data results, potentially precluding the generalizability of findings to a larger population. Additionally, as the students were of a fairly advanced level—American university students with high TOEFL scores (around 550) and 5+ years experience living in the U.S.—there is some question as to how much we may be able to transfer the results to other types of ESL situations, especially those involving lower proficiency students. Furthermore, in terms of Stanley’s conclusions, while it was encouraging to see that the coached group employed a greater use of strategies, devoting 7 hours to peer review preparation is probably beyond what is practicable by most language teachers, so it may have been more effective to limit the time devoted to coaching to that which could be completed in 1-2 class sessions as this could be more manageably applied to other contexts. Finally, it is unfortunate that very brief space was given to discussion of student revisions, with almost no conclusions drawn from this except that the coached group made more revisions based on evaluator comments. Mention of whether these comments were valid and whether revisions made which were based on them led to an overall improvement in the work would have been insightful in terms of connecting peer review with better student writing. This was not the central focus of the study, however.

As an earlier example of ESL research into peer review considerations,
Stanley’s research exemplifies the benefits that extended planning can bring to allowing students to comfortably participate in these activities. There are also strong clues as to the role that training can play in helping them to become better evaluators—a notion borne out in later studies.

Cassia Mendoça and Karen Johnson (1994)

Summary

This study of peer response and its effect on ESL students’ writing was undertaken with some of the same goals in mind as the Stanley study. Titled “Peer review negotiations: Revision activities in ESL writing instruction,” the authors were interested in what types of negotiations occurred during peer review sessions (similar to Stanley’s study from 2 years prior) though they also sought to find how these sessions ultimately impacted later revisions of written work. In addition, the researchers were also interested in gathering student feedback on the activity itself in the form of post-interviews.

The participants of this study were 12 graduate students at a large university, presumably in the U.S., who were taking a course with the goal of preparing them for research-oriented writing in English. Three types of data were collected: tape recordings of peer review sessions, student drafts of written assignments, and tape-recorded post-interviews. There was no mention of coaching students before the sessions; only guided questions such as, “What is the main idea of your partner’s paper?” and “What suggestions could you give your partner?” were used to help start discussion. The peer sessions were later analyzed using “analytic induction procedures” based on an earlier study by Bogdan and Biklen (1992). This coding of student responses is similar in some ways to that created by Stanley (1992), although there was no distinction made between evaluator and writer comments. Finally, the post-interview recordings were checked
by the researchers in order to determine the students’ attitudes towards the activity.

In their results section, Mendoça and Johnson display the frequency of occurrence of different types of “negotiations” used by the students during their peer review sessions. Negotiations were divided into five categories: Questions, Explanations, Restatements, Suggestions, and Grammar corrections. Each of these categories was explained in-depth and illustrated through excerpts of dialogue from students’ tape-recorded sessions. A distinction was made between the number of writer-generated and reviewer-generated comments, with reviewer comments dominating in most categories. In terms of revisions on subsequent drafts, the researchers found 21 instances in all 12 students’ revised compositions which were directly related to peer discussions. The authors gave several examples of these revisions and again included excerpts from the recorded dialogues in cases where these were based directly on peer comments. Finally, summaries of the post-interviews (conducted by the researchers) were included with 10 of the 12 students reacting favorably to the exercise.

From this data, Mendoça and Johnson interpreted their findings as “confirmation” that “peer reviews force L2 students to exercise their thinking as opposed to passively receiving information from the teacher” (p. 765). They also felt that the activity improved students’ ability to “express and negotiate their ideas” and “develop a sense of audience” (p. 766).

Analysis

As with the earlier Stanley (1992) study, we find in this research results which favor the use of peer review activities, and in this case the voice of the students also lends its support. Multiple instances of learners incorporating peer suggestions into their subsequent written drafts are the type of evidence that can be seen as concrete substantiation of the process’ benefits. These could even be seen as providing a route towards greater
student autonomy and critical reading skills and a lessening of the teacher’s role, which tends to dominate traditional writing classes.

Yet while the conclusions drawn by the researchers can be taken to indicate that peer review is an important component of the writing process, certain limitations in the study would seem to qualify the results. First, and perhaps foremost, and as mentioned by the co-authors in their conclusion, the small sample size used makes it difficult to draw generalizable conclusions from such a limited collection of data. Using writing and speaking samples from only 12 students forced the researchers to closely examine, and perhaps “over-analyze” individual examples of peer review comments and writing revision, thus making each one seem more significant than it perhaps is in reality. Some dialogue excerpts were even repeated at different points in the article in order to demonstrate different processes, thus indicating the limitations of having such a small data set. In all, 40 revisions were made to the 12 students’ writing, and of these, only 21 were considered directly attributable to comments made in the peer sessions. Perhaps it would have helped to collect data from a second group for comparison as Stanley did. Also, it seems Stanley conducted far more peer review sessions than were done by Mendoça and Johnson, which could explain her more definitive results.

Another consideration is that the conclusions which were arrived at after carrying out a study are limited in scope to the participants involved in the study. The results are encouraging for instructors with higher-level students, which is logical considering learners often have fewer inhibitions to speak out than those at lower levels. However, caution must be taken when applying these activities to contexts involving beginning-to-intermediate level or mixed proficiency students.

In addition to these points, there are a few smaller issues that may also affect the decisiveness of the conclusions. It was surprising that the students were given no coaching at all as to how to effectively evaluate their peers.
Instead, they were asked to begin their peer sessions with some rather vaguely-worded guided questions. Perhaps some more time spent acclimating them to the process would have resulted in more peer suggestions.

Despite these concerns, Mendoça and Johnson’s results seem to bear out the benefits of incorporating peer review into a writing curriculum. From their research we can understand that following up activities with assessment of how many suggestions were incorporated into final drafts, as well as asking for student feedback to the peer review process, can be important elements to research in this area.

*Olga Villamil and Maria De Guerrero (1998)*

**Summary**

This study, “Assessing the impact of peer revision on L2 writing,” was done in 1998 at a university in Puerto Rico. It sought to add a new dimension to existing L2 peer review research by examining the impact of such an activity on written revisions in two different rhetorical modes: narrative and persuasive writing. Fourteen Spanish-speaking participants were chosen to determine how and whether peer review suggestions were incorporated into writers’ final drafts, as well as how “troublesources”—defined as “perceived or potential problems…in the text which were sources of revision”—were “revised according to different language aspects [i.e., content, organization, grammar, vocabulary, mechanics]” (p.497).

The 14 participants in the study were drawn from a larger pool of students who had taken part in two previous studies concerning peer review done by Villamil and De Guerrero (1994, 1996). After being exposed to the two rhetorical modes under study—narration and persuasion—students were introduced to peer review processes by being shown essays composed by previous class participants and were then asked to suggest revisions
based on guided questions asking about the content and organization of the essays. Revisions were then discussed by the whole class so that students were adequately prepared for the task of peer review. They then went on to compose their first drafts in the rhetorical mode being studied and then worked in pairs to make evaluative comments about each other’s work. These discussions were tape-recorded and later transcribed.

In their analysis, Villamil and De Guerrero later divided the troublesources into five language areas—content, organization, grammar, vocabulary, and mechanics. The researchers found that the majority (74%) of troublesources discussed and revised during the peer sessions made it into the final drafts. Additionally, 39% of all revisions in the final drafts were self-revisions not discussed during peer sessions. Looking at the results for the two different rhetorical modes, discussions on narrative writing generated nearly the same number of revisions as persuasive writing. As for types of corrections made, grammar revisions predominated and organization was least dominant in each type of writing. Both modes also saw benefits in terms of increased length of final drafts after peer revision, although narrative compositions saw more lengthening (68 versus 38 words added on average). A final analysis of data from re-written drafts noted that 7% of all revisions were false repairs—incorrectly revised parts of the text. Roughly half of these derived from peer discussions while the rest were a result of self-repair.

Villamil and De Guerrero noted that, similar to Mendoça and Johnson’s study (1994), not all suggestions were incorporated into final drafts, indicating that the nature of peer interactions might determine how many revisions are made based on peers’ comments. For instance, friendlier pair relationships may result in more suggestions being incorporated into later drafts. Furthermore, the authors considered the large number of post-session self-revisions as a sign of “self-regulatory behavior”—independent performance being influenced by peer assistance.
Continuing the discussion, the researchers reiterate the fact that the most significant trends in students’ revised drafts were seen in grammatical correction (perhaps reflecting that most of the comments made during the peer sessions were grammatical in nature) and text lengthening (which indicate a rethinking of the writing based upon peer comments). The authors then conclude by summarizing the findings mentioned above and restating the various benefits that peer review activities can have on students for both skills-oriented and social-psychological aspects of learning a second language.

Analysis

Building on previous research into the area of peer review effectiveness, this study in several ways strengthens the argument for using such activities in the language classroom. First, the authors included a quite thorough review of the literature which referred to nearly every major L2 peer review study made up to that point. Where there were gaps, they attempted to fill these by looking at existing L1 research. The inclusion of social interaction studies by Vygotsky (1978) also helped to inform the reader as to psychological factors that might be involved.

Other strengths of the study were also apparent. First, the system for coding students’ negotiations was well-conceived, and the inclusion of draft and dialogue examples helped to put these into context. Additionally, the results of this research were well-explained with the use multiple tables which helped to illustrate students’ tendencies in both comments and written revisions. Several variables were examined, such as draft lengths and the number of revisions per 100 words, which were left out of the other studies. Finally, the extended discussion section looked at nearly every aspect of the results and attempted to put these into a broader perspective concerning the implications for our understanding of language acquisition.

This latter point deserves some special mention. Villamil and De
Guerrero make a serious attempt in this study to ground the justification for their research in the social-constructivist theories of Lev Vygotsky. This is an important consideration often overlooked by other researchers into the peer review process. Vygotsky’s ideas, especially his notion of Zones of Proximal Development (1978, 1986), underlie many of today’s socially-oriented approaches to education, and the increase in the usage of peer review activities is due at least in part to the spread of his ideas. Thus, an examination into the effectiveness of such activities could also be seen, from a broader perspective, as putting Vygotsky’s ideas to the test. The conscientious acknowledgement in the Villamil and De Guerrero study of their theoretical underpinnings is a model for future research in this field.

Despite the thorough nature of their analysis, however, there were some limitations in design which should be noted. As with the other studies, only a small number of participants were used—14 in this case—who engaged in the two peer review sessions. Additionally, the students being employed had already been involved in two previous peer review studies carried out by the same research team within the prior four years. Although it is not uncommon for researchers to use data from the same participants in multiple studies, care in such cases should be taken when interpreting the results as repeating the same process runs the risk of the students becoming too familiar with the procedures and expectations involved, and it can make the findings more difficult to apply to non-initiated students in other ESL contexts.

Elaine DiGiovanni and Giriga Nagaswami (2001)

Summary

The authors of this study, titled “Online peer review: An alternative to face-to-face,” took a very different approach to peer review than was conducted in the previous articles in terms of procedure and the research
questions asked. DiGiovanni and Nagaswami were interested in comparing results of peer reviews done by students in a traditional face-to-face setup (FFPR) with those done online using networked computers (OLPR). They hoped to discover the different types of negotiations used by students in both scenarios as well as how much training would be necessary for the online portion of the exercise. To this end, 32 U.S. community college students from two ESL writing classes were selected. Students were grouped into pairs depending on differences in first language, English proficiency, and computer literacy.

The participants in the study received training in both FFPR and OLPR through role-play modeling by the teachers, whole class discussion of these role plays, introduction of teacher-created peer review guidelines, and use of the software to be employed for the OLPR sessions. As the authors were more interested in evaluation related to content and organization, they discouraged the students from being “grammar editors.” Early in the course, several sessions of traditional peer review were done by the students in pairs using printed guidelines. Initial online assignments included collaborative brainstorming and composing activities while later activities involved commenting on first drafts of assigned compositions using guided questions. Finally, all students were given questionnaires at the end of the course which asked about their impressions of the peer review activities.

Of the 32 students involved in the study, 10 were chosen for data collection and analysis of the peer review sessions. As with the previous studies, tape recordings were made of face-to-face sessions while comments elicited during online sessions were later printed out for examination. For analysis of the data, DiGiovanni and Nagaswami divided student comments into categories based upon those used in the earlier Mendoza and Johnson study (1994).

The results show that similar types of peer review comments were used in roughly equal measure in both contexts, but with the greater number overall
being made in the face-to-face scenario (381 versus 102). The most utilized types of negotiations included requests for information, explanations of unclear points, agreeing or disagreeing with ideas, and restatements. Unlike Mendoça and Johnson (1994), DiGiovanni and Nagaswami made no distinction between writer and evaluator-generated comments. Questionnaire results showed that the vast majority of students (>90%) responded positively to the experience and that a significant majority of responders preferred the online context to traditional face-to-face sessions. In their discussion of the activity, DiGiovanni and Nagaswami found OLPR to have several advantages for students. These included students staying more on task, the teacher having a greater ability to monitor this type of interaction and guide students who appear to be making overly general comments, and also that students could refer to printouts of their discussions instead of relying on memory to make revisions of their written work.

**Analysis**

Although this study was done briefly and used only a small number of participants (10 for both types of peer review sessions), the alternative approach to peer review considered here is intriguing. First, the researchers looked at important aspects related to how students comment on each other’s work in two very different contexts. Of interest is the fact that nearly four times more comments were made during the face-to-face session than during the online sessions. Might this be because students felt less comfortable critiquing one another in the online setting? Further research could help to illuminate this issue. Nonetheless, students on the whole reacted more favorably to the OLPR component, which is encouraging when we consider the greater role technology is already taking in some language classrooms. Another positive distinction from the previous studies is the fact that two different levels of students were used for the study. It would be interesting to find out which group felt they had derived more
benefit from using OLPR as opposed to FFPR.

Despite its promise, a few factors in the study could be seen as being limiting the veracity of its conclusions. First, it was not stated why data was collected from only 10 of the 32 students in the classes, but this undoubtedly affected the strengths of their conclusions based on the small data set. Also, it would have been helpful to know how many peer comments and suggestions were actually incorporated into subsequent student drafts, as was done in earlier studies. Furthermore, the researchers relied heavily on student feedback to the activities to justify the use of OLPR over FFPR as well as mentioning teacher benefits for engaging in this type of activity. Stronger support would most likely have been indicated by the number of follow-up revisions made by the students based on peer session comments.

Despite these drawbacks, this study is one of the first to examine the effect that computer use can have in the peer review process—a role which is seen as increasingly important in many of today’s multi-modal writing classes which are heavily dependent on interactive technology. As word processing and other software continues to incorporate a wider variety of editing and commenting features for both writers and readers, teachers in L2 contexts who take advantage of these to enhance the writing process for their students will be helping them to stretch the boundaries of traditional essay writing. Peer review can be seen to derive obvious benefits when matched with editing technology, and further research which probes its possibilities as the present study does will no doubt be seen more frequently in the literature in years to come.

*Hui-Tzu Min* (2005)

**Summary**

The final and most recent study being examined is “Training students to
become successful peer reviewers,” written by Min in 2005. The researcher, a professor at a Taiwanese university, investigates the effect of peer review within a different environment from the previous studies—namely a writing course in a definitively foreign language (as opposed to second language) context. Min begins by noting the lack of EFL studies in this area, despite the fact that cultural factors may inhibit such students from critically evaluating their peers’ work more than those in ESL environments. With this issue in mind, he set out to determine the effect of extensive training on student peer review output.

Participants in the study were 18 Taiwanese students taking a composition course at a university in Taiwan. The first stage involved engaging the students in a peer review activity before receiving any training in order to establish a baseline to which later data could be compared. In response to reading each other’s essays, Min found their comments at this point to be overly brief and perfunctory in nature; in other words, of minimal use to the original writer. After this, the next phase included 4 hours of in-class training—through modeling by the teacher—followed by student-teacher conferences which included tailored assistance in making useful comments on their peer’s compositions. From there, students then had two hours to read and respond to two their classmates’ short essays (~250 words on average).

For his analysis, Min used both quantitative and qualitative methods. The former involved the use of statistical t-tests comparing the number of comments, words, and global and local issues found in the essays before and after the peer review training. The latter drew from student journal entries, which included reactions to the activities. For the t-tests, a significant difference was found when comparing the mean number of words and comments used by the peer reviewers as well as how many “steps” were used in feedback. These steps included multifaceted suggestions which served functions such as clarifying the writer’s
intentions, explaining the nature of the problem, giving examples as suggestions, etc. Finally, the researcher also discovered that the mean number of global suggestions (though not local suggestions) had increased after the training.

In his examination of student journals, Min found reaction to the peer review activities (after training) to be decidedly positive. Several areas of improvement were mentioned including: the use of “steps” to bring focus to the evaluators’ comments, increased self-monitoring when producing subsequent compositions, and a gain in confidence as readers. Many also felt that the training had improved their compositions overall as it helped them to be more conscientious as writers.

The discussion section of the study includes a number of conclusions made from analyzing the data. In addition to the use of extensive training leading to significant benefits in terms of evaluative skill, Min also mentions the necessity of peer reviewers attempting to clarify their impression of the original writers’ intentions before proceeding with suggestive comments. He claims that doing so will help the writer to better understand the perspective of the reader. There is also a brief discussion of the social-constructivist benefits of peer review as per Vygostky’s theories.

**Analysis**

The context and research of this study represent, in many ways, a significant diversion from the articles previously examined. It is refreshing to see the focus on students in an EFL environment, as they undoubtedly face more challenges when engaging in peer review activities than do participants in English-speaking circumstances. As Min mentions, using students with a shared L1 background may lead them to make assumptions about each others writing that are not necessarily true in reality (hence his emphasis on evaluators clarifying what they think the writers’ intentions are in their comments). One can imagine that certain cultural factors may come
into play as well, especially when direct peer criticism may be avoided by students in certain parts of the world. Unfortunately, Min chose not to focus on such contextual inhibitions in his examination.

Also encouraging was the effect of extensive training on students whose peer review comments were initially reticent and unhelpful. In this way, the current study bears similarity to the findings Stanley (1992) made in her research; so perhaps successful peer evaluation depends less on cultural considerations than on proper coaching. Nonetheless, the fact that, once again, so few participants were involved casts some measure of doubt on the generalizability of the findings. Further study involving 50 or more students would most likely be more enlightening.

The use of basic statistical analysis also set this research apart from the others examined here. Though some might argue that comparing means of word counts and grammatical errors runs the risk of over-scrutinizing one’s data, in this case the researcher used these results in conjunction with other, more qualitative sources (journal entries) to arrive at his conclusions. In following this mixed-methods approach, statistical analysis can be seen to enhance the strength of one’s findings. (However, it should be noted that Min made a common statistical error when running multiple t-tests in one of his analyses by not adjusting the significance level downward.) Although some researchers may hesitate to include statistical procedures in a study on such a qualitatively oriented area as L2 writing, there are cases involving such components as comparison of pre and post-treatment differences where they can be quite useful. As long as care is taken to consider such results as part of a larger set of data, as Min has done here, more writing research can benefit from the objectivity of quantitative analysis.

**Discussion**

The substantial body of research examining the implications of conducting peer review activities in various L2 contexts is an indication of
the increasing popularity of social-constructivist oriented curriculums developed by language teachers around the world. It is encouraging that all of the L2 studies considered here found benefits of some kind in utilizing such activities. Moreover, differences in research questions and approaches resulted in a variety of perspectives from study to study. Whereas Stanley (1992) and Min (2005) were mostly concerned with the effect of extensive training on the effectiveness of peer review sessions, Mendoça and Johnson (1994) and Villamil and De Guerrero (1998) chose to give more attention to how many evaluator suggestions were incorporated into students’ revisions, and DiGiovanni and Nagaswami (2001) looked at alternate contexts through which peers can undertake to critique each other’s work. Furthermore, some researchers focused on putting students into pairs while others were more interested in following peer review groups of 3-5 participants. Despite these differences, similarities did exist from study to study. Several of the researchers tape-recorded and transcribed review sessions, and everyone analyzed revision comments using some type of coding system. Having common practices like these provides a rudimentary unity which helps one to better understand the studies in relation to one another.

While each study brought new insight into the overall research area being examined, some reservations must also be voiced. First and foremost is the ongoing concern regarding the small number of participants used. Fewer participants means less data, which makes it more difficult to draw definitive conclusions about any one aspect of peer review research. The fact that all of the authors determined that these types of activities can play an important role in ESL writing classes was heartening and speaks well as an overall endorsement for using peer reviews, but it would have been better to have had a large amount of data to support these conclusions. Another fact which must be considered is the preponderance of research done in ESL as opposed to EFL contexts. Most studies to date have involved advanced-level students at U.S. universities, a reality which is only
now beginning to change as more research from Asia make its way into the literature. Students with extensive experience living in English-speaking environments will have attained a level of proficiency that would most likely put them closer to L1 participants than L2 learners of much lower ability living in non-English speaking environments. For greater balance, more research involving EFL learners and/or learners with mixed-proficiencies, would be a welcome attraction to the literature. Especially interesting would be the inclusion of cultural attitudes and differences as a component of such studies.

Finally, the addition of a qualitative evaluation of the students’ writing by the researcher/teacher could also be an avenue for future research. While there is benefit in counting the number of revisions made in student drafts, it would be helpful to know whether there were any qualitative improvements in students’ writing (according to a teacher or researcher) after peer review was performed. This would bring a more subjective aspect into the studies, but it could be a useful indicator when considered with the other results of data analysis. After all, if students’ writing is not being improved, one has to question whether a significant amount of class time should even be devoted to peer review sessions.

In summary, the following aspects from the currently discussed studies have benefited research into peer review and merit consideration for any future research:

1. The peer review process is a natural outcome of social-cognitive theories of learning, and investigation into this background can help the researcher to design a study within this theoretical framework.

2. Acclimating the students to the peer review process through training and modeling tends to increase the specificity and usefulness of peer comments.
3. Use of an established coding system when analyzing peer suggestions can aid analysis and also provides a frame of reference for other researchers to better comprehend the results.

4. Inclusion of multiple types of data in a study, such as quantitative and qualitative information, helps to provide a more holistic understanding of student engagement in and responses to the peer review process.

5. Research involving multiple modes of peer review activities, such as handwritten, face-to-face, and computer-aided processes, is an area of emerging interest.

6. Students learning writing in EFL contexts may benefit from less direct peer review approaches than those in ESL environments.

In addition to these conclusions, certain limitations in these studies could also be addressed in future research:

1. A larger number of participants than those found in most current research, more than 30 for example, could yield data which can be more concretely analyzed, especially using statistical methods. This can also improve generalizability of the results to other contexts.

2. Studies involving learners at mixed proficiency levels, particularly those who lie at or below intermediate English ability, would help to broaden the scope of peer review research. It is important to understand whether such activities benefit low-level learners in addition to advanced learners.

3. Qualitative assessment by the teacher and/or researcher of students’ revised drafts after peer review has taken place may further illuminate the outcomes of the process.
While it would perhaps be unduly ambitious for a researcher to attempt to address all of these considerations within the scope of one study, several of them could easily be incorporated into an existing design. As an EFL writing teacher and researcher, I have approached my recent research with these issues in mind, and what follows is a brief summary of a study in progress.

Neff(2006): Results from an ongoing study

The primary aim of this research has been to determine how EFL university students respond to two forms of peer review activities—written and oral feedback. Forty-eight participants of varying levels of proficiency in two separate classes are involved in the study. I have been curious to see whether students would respond more positively to evaluation involving face-to-face discussion or written notes. To this end, both approaches were introduced to each group in alternating order in a so-called counterbalance design. Once they had finished the activities and re-written their compositions based on the peer suggestions, the students were then given questionnaires asking them to count the number and type of suggestions made as well as their overall reaction to the two approaches.

Preliminary results, which have been analyzed using both statistical and coding methods, indicate a definitively positive response to the activities as a whole, with a slight bias towards engaging in face-to-face peer review. Many students felt that they benefited from reading their peers’ work critically and were also satisfied to have their own work taken seriously. Most of their peers’ suggestions were incorporated into later drafts, though I have not determined how many of these suggestions were correct (a consideration which deserves more attention in peer review research). Also, no qualitative assessment of the students’ work has been matched to drafts written after peer review.
Perhaps most surprising to me as a researcher was the fact that many students enjoyed, and even preferred, face-to-face peer review, despite any reticence they may have had in critiquing their classmates this way. By far the largest number of comments made to support this approach mentioned the ability to ask for clarification of evaluator suggestions—a facet which was missing from the written approach. I also feel that properly training the students for at least one hour in peer review methods helped them to become used to the concept and understand its benefits. Once again, there appear, to be clear advantages derived in taking time to explain how to make the right type of comments in a considerate fashion.

Considering how much focus has been given thus far to the peer review process in ESL and EFL contexts, there still appears to be a number of areas and variables that remain largely unexplored. Examination of these could lead to better comprehension of the benefits gained from having students critique each other’s written output. The future would seem to hold significant promise for researchers looking to explore new aspects of conducting peer review in the language classroom.

References


概要

近年、英作文の分野において、「ピア・レビュー（相互評価）」の必要性が強調されている。ピア・レビューの概念のもとで学生同士が英作文を評価し合い、コメントしあうことは効果的であり、必須であると多くの外国語教師が認識している。特に、ピア・レビューが提唱した「社会の中で交流しあうこと」の大切さが論文として発表（ピア・レビュー）されて以来、ピア・レビューは教授法の一要素として定着しているが、その効果は定性的に検討されるべきである。ピア・レビューは英語を第二外国語として指導する教師らにより英作文カリキュラムに採用され、多くの研究が行われた。本論文ではそれらの研究のうち、いくつかを取り上げ、英語学習者にとって
のピア・レビューの重要性について検討した。同時に、それらの研究の長所、短所を分析するとともに、様々なタイプのピア・レビューについて考察した。最後に、今後の課題を述べ、まとめとした。