International Journal for the Scholarship of Teaching and Learning

Volume 5 | Number 2 Article 14

1-2011

How Do Students of Diverse Achievement Levels Benefit from Peer Assessment?

Lan Li
Bowling Green State University, lli@bgsu.edu

How Do Students of Diverse Achievement Levels Benefit from Peer Assessment?

Abstract
Although the potential of peer assessment activities in promoting student learning and fostering student cognitive development has been widely studied and well documented across the world, it is unclear how peer assessment may benefit students of diverse achievement levels. This study examined this issue via a mixed methodology approach that combined students' project scores and survey responses with a qualitative interview technique. Findings of this study suggested that peer assessment activities employed in this study had a differentiated impact on students' learning and perceptions. Students in early learning development stages showed more learning gains than high achieving students in terms of point increase in their grades. Nevertheless, students across diverse achievement levels generally held positive attitudes towards their peer assessment experience.

Keywords
Peer assessment, Collaborative learning, Formative assessment, Peer interaction

How Do Students of Diverse Achievement Levels Benefit from Peer Assessment?

Lan Li
Bowling Green State University
Bowling Green, Ohio, USA
lli@bgsu.edu

Abstract

Although the potential of peer assessment activities in promoting student learning and fostering student cognitive development has been widely studied and well documented across the world, it is unclear how peer assessment may benefit students of diverse achievement levels. This study examined this issue via a mixed methodology approach that combined students' project scores and survey responses with a qualitative interview technique. Findings of this study suggested that peer assessment activities employed in this study had a differentiated impact on students' learning and perceptions. Students in early learning development stages showed more learning gains than high achieving students in terms of point increase in their grades. Nevertheless, students across diverse achievement levels generally held positive attitudes towards their peer assessment experience.

Key Words: peer assessment; collaborative learning; formative assessment; peer interaction

Background

Peer assessment is the process of students evaluating the work of peers using applicable marking criteria (Falchikov, 2001). Peer assessment activities can be categorized as either formative or summative. Formative peer assessment aims to cultivate student learning and usually involves students in reviewing peers' work, with the purpose of assisting peers in further developing their projects. In such a process, students play both roles of assessors (rating and commenting upon peers' work) and assessees (viewing and acting upon feedback). On the other hand, summative peer assessment emphasizes accountability. In summative peer assessment, students play only the assessor's role by judging the quality of peers' work and assigning quantitative marks or grades (Topping, Smith, Swanson & Elliot, 2000; Xiao & Lucking, 2008).

Peer assessment, both formative and summative, may foster student cognitive development and active learning in various ways. Multiple potential benefits have been reported from previous studies: enhanced learning outcomes (e.g., Li & Steckelberg, 2005; Pope, 2001), deeper understanding of quality performance (Falchikov, 1995; Dochy, Segers, & Sluijsmans, 1999; Partri, 2002), increased sense of autonomy and motivation (Brown, 2004; Hiltz & Wellman, 1997; Pope, 2001), adequate and timely feedback (Gibbs, 1999), and enhanced sense of responsibility (Somervell, 1993). Cheng and Warren (1999) further portray peer assessment strategies as a reflection on "what learning has taken place and how" (p. 301). Topping (1998), based on an extensive review of 109 peer assessment studies conducted across the world, summarizes that the peer assessment process benefits both assessors and assessees in multiple ways, such as increased time on task, focused attention on critical elements of quality work, and greater sense of accountability and responsibility. Furthermore, he posits that these cognitive gains "might accrue before,

https://doi.org/10.20429/ijsotl.2011.050214

during and after" the peer assessment process (p. 256).

Statement of Problem

Despite the great potential of peer assessment, it still remains unknown what makes peer assessment effective (Van Zundert, Sluijsmans & Van Merrie¨nboer, 2010). Specifically, it is unclear in regards to how peer assessment benefits students on various learning levels. Previous studies in this area are scarce. These few studies have raised and discussed this "how students benefit" issue as a peripheral finding. For example, Davies (2000) introduced a Computerized Peer Assessment system (CPA) as a part of the assessment process in a computer studies class. Participants wrote and submitted a report, and then took a multiple choice pre-test in the area of computer, communication and networks. Afterwards, each participant assessed ten peers' reports in the CPA system. Upon completing the peer assessment process, participants took a multiple choice post-test to evaluate their knowledge on the designated topic. Although the purpose of Davies' study was to report the functionality of the CPA system and assess its effectiveness from the perspective of students' perceptions, analysis of students' reports and tests indicated that certain groups of students (grouped by grades of their initial reports) benefited differently from this peer assessment process. Particularly, low-achieving students statistically benefited most from their participation in the CPA peer assessment activities, while the least benefits were shown for high-achieving students. This interesting side finding has been partially supported by some later studies.

In one of their peer assessment studies, Li and Steckelberg (2004) randomly assigned students into two groups: experimental and control. After students completed their projects, the experimental group underwent peer assessment to improve their projects, while the control group continued working on revising their projects by themselves during the same period of time. The students' initial (before peer assessment) and final (after peer assessment) versions of projects were independently evaluated and compared by a researcher and a trained grader. The authors found the variability of the final scores in the experimental group was substantially smaller than that of the control group, which might suggest that peer assessment had a differential impact on students of diverse achievement level. This hypothesis was echoed by one of their later studies that explored student perceptions toward peer assessment, as one student commented, "Sometimes peer assessment isn't helpful if you already did a good job" (Li, Steckelberg & Srinivasan, 2009, p.143). These interesting results have drawn the researcher's attention to the important yet neglected question of how students benefit from peer assessment activities.

In view of this, the current study was designed to determine the relationship between students' diverse achievement levels and potential benefits of peer assessment and to explore perceptions of students towards peer assessment. The research questions were: 1) In a structured peer assessment process, what were the learning gains of students of diverse achievement levels, as represented by their different levels of prior knowledge of a designated topic? 2) What was the experience of students of diverse achievement levels in this peer assessment process?

Methodology

A mixed methodology design was utilized in this study, as the mixed method research approach allows for a triangulation of data (Johnson & Onwuegbuzie, 2004), which offers an in-depth perspective that a single quantitative or qualitative design may lack. Specifically,

a pre-post design (quantitative) was triangulated with survey responses (quantitative and qualitative) and interviews (qualitative) in this study.

Participants

Participants included twenty-one graduate students enrolled in two classroom technology classes at a Midwestern university in the United States. The demographic survey indicated that most participants were European American (n=18), female (14), and full-time students (15) majoring in varied education-related master programs such as Curriculum and Teaching, Special Education, and Cross-cultural Studies. Most participants had no prior experience with formal and structured peer assessment (19) or WebQuest creation (18).

Context and Procedure

All participants followed the same peer assessment process. First, students were informed that they would participate in a peer assessment activity. Students then completed a demographic survey that included information such as gender, age, race, major, information about previous experience with peer assessment, and prerequisite knowledge of the content area (WebQuest). Afterwards, the specific peer assessment procedure was presented to the students, and they were encouraged to ask questions related to peer assessment. The purposes of this step were to provide an opportunity for students to learn the potential of peer assessment and motivate them to fully engage in this activity.

Second, students learned how to create a quality WebQuest and present its content in a word processing document. Creation of WebQuests in this study took one week. Students learned the concept and critical features in class, and composed and submitted the content of their own WebQuests in the following week. A WebQuest, as defined by Bernie Dodge, is an Internet-based lesson that utilizes web resources and provides practice for students' higher order thinking skills (Dodge, 1995). A WebQuest provides a series of scaffolding activities to aid student learning. The five critical features of WebQuests include Introduction, Task, process, Evaluation and Conclusion (MacGregor & Lou, 2004). Introduction provides background information, grabs learners' attention and orients them to the incoming tasks. Task describes end products of WebQuest activities. In the Process section, step-by-step instruction, scaffolding activities, and resources are provided to guide learners to complete tasks. Evaluation illustrates how learners' performance in WebQuests would be assessed, while Conclusion summarizes learners' experience and encourage reflections. Since its creation in early 1995, student and teacher design of quality WebQuests has often been considered an effective training strategy in teacher preparation programs (Li, Liu & Steckelberg, 2009).

Third, students viewed and discussed marking criteria in class. These marking criteria were initially developed in 2004 in accordance with a few WebQuest resources (e.g. Bellofatto, Bohl, Casey, Krill & Dodge, 2001; March, 2011; Martin, 1999) and have been modified and revised according to students' and instructors' feedback since then. The marking criteria consisted of eight categories with 40 total points, covering the five critical sections of the WebQuest. Training was provided to enable students to understand the critical features of WebQuests and to practice their assessment skills. Two examples of WebQuest projects were provided in class, and students were requested to grade these projects based on the marking criteria. The teacher's and students' ratings of these example projects were discussed and compared to enhance students' understanding of the content and the marking criteria, as well as how their performance would be evaluated.

Fourth, students composed and uploaded their own WebQuest projects in a Word document to a Discussion Board forum in Blackboard. Blackboard is a course management system that

provides an integrated set of web-based tools. Like typical bulletin boards, the Discussion Board in Blackboard allows students to post and respond to threads asynchronously.

Fifth, the instructor randomly assigned each student two peers' projects to review. As a result, each student reviewed two peers' projects, and at the same time two peers reviewed his project. Students rated and commented upon peers' projects based on the marking criteria previously discussed. Specifically, the expectation was clearly communicated that each reviewer should not only rate each peer's performance but also provide constructive suggestions for further improvement.

Sixth, each student was required to revise and improve his WebQuest based on experience in this activity and the peer comments received. In this step, students were specifically instructed that peer feedback may vary significantly in quality. Therefore, students needed to judge the quality of peer feedback received and decide which suggestions to adopt.

In the seventh and last step, students submitted their revised WebQuests to the instructor through Blackboard. Students then completed the post-assessment survey. The survey explored students' perceptions regarding the peer assessment activities and the quality of peer feedback (See Data Collection and Analysis for details).

Based on the students' initial knowledge of the content area (WebQuest), students were divided into 3 achievement groups. Six students (two in each group) were purposefully selected to participate in the post-assessment interviews (See Data Collection and Analysis for details).

Data Collection and Analysis

Both quantitative and qualitative data were collected. The quantitative data included scores of students' initial and revised versions of WebQuest projects and students' survey responses. The qualitative data included student responses to open-ended questions in the post-assessment survey and interviews regarding students' overall perceptions of the peer assessment they experienced.

Quality of students' initial and revised WebQuests (quantitative)

The quality of the initial and final versions of WebQuest projects represents students' content knowledge before and after peer assessment, respectively. Following the same marking criteria that students used in peer assessment, the researcher and an independent grader blindly judged the quality of the initial and final versions of the students' WebQuest projects to ensure grading consistency. A panel of content experts determined the face validity of the marking criteria. Three instructors teaching the same course had reviewed these marking criteria and agreed that these criteria could accurately measure quality of students' WebQuest projects. The inter-rater reliability between the researcher and the grader was satisfactory with 0.95 for the grading of the initial WebQuest projects and 0.88 for the grading of the final WebQuest projects. Data analysis employed the average scores of the researcher and the grader. Based on the students' performance on their initial version of WebQuest projects and the marking criteria, the twenty-one students were assigned to three groups representing three achievement levels: Beginning (5 students), Developing (10 students) and Advanced (6 students). The average point increase in each group was calculated.

Post-assessment survey (quantitative & qualitative)

The post-assessment survey asked if students were satisfied with the quality of their own WebQuest projects, and how students thought peer assessment activities influenced their

WebQuests. Further, students were prompted to rate the quality of peer feedback they had received in the categories of Very Good, Good, Barely Acceptable, Poor and Very Poor and to justify their ratings. Students were then asked to rate seven 5-point Likert Scale statements regarding their perceived value of peer assessment activities. These statements were adapted from previous studies (Li, 2007; Lin, Liu & Yuan, 2002). Internal consistency reliability was tested in a previous study and was 0.92 (Li, 2007). Survey data were coded in SPSS (Statistical Package for the Social Sciences) to calculate frequency.

Interviews (qualitative)

More qualitative data were collected by interviewing students after the post-assessment survey. A theoretical sampling technique (Chenitz & Swanson, 1986) was employed to select an appropriate interview sample. In each achievement group (Beginning, Developing and Advanced), the two students who were identified as having the most and the least positive attitudes toward peer assessment (based on survey responses) were invited for the interview: Beth and Ben from the Beginning Group, Dana and Darcy from the Developing Group, and Allison and Aaron from the Advanced Group (all pseudonyms).

The researcher used a semi-structured format for the interview. Each interview was approximately 20 minutes in duration. Students were asked questions regarding their confidence level of conducting peer assessment, their thoughts of the quality of the projects reviewed, and how they thought the peer assessment process influenced their learning and project quality. The following are some sample questions asked in the interview: 1) What did you think of the quality of the project(s) that you reviewed?; 2) What did you think of the quality of the peer feedback that you received?; 3) How do you think the peer assessment process has impacted your learning/project?

All interviews were conducted, audio taped, and transcribed, either by the researcher or a trained research assistant. Transcripts were coded and analyzed for common meanings and themes, using two levels of coding. In the first level (Initial Coding), transcribed data were read through to identify meaningful patterns, and labels/codes were added to sort and assign meaning to text. In the second level (Focused Coding), labels/codes were reviewed, eliminated, combined or subdivided to identify repeating ideas and themes. Figure 1 shows an example of how a small passage of a participant's interview transcript was coded.

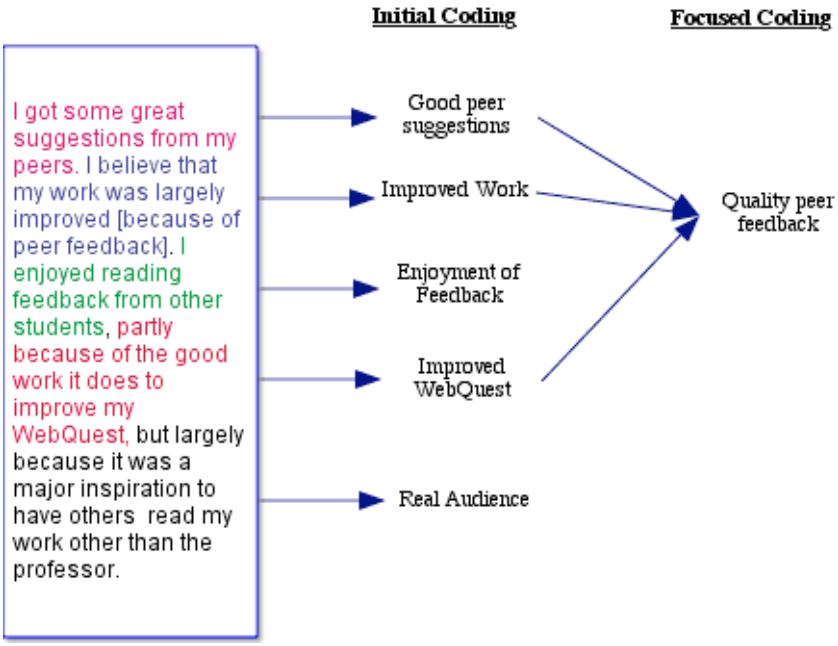

Figure 1. Coding Tree Demonstrating How Interview Transcript Was Coded

Interview transcripts were coded individually by the researcher and an independent trained coder. Labels and themes generated from these two coders were compared to identify agreement and disagreement. The formula to test coding reliability suggested by Miles and Huberman (1994) was utilized:

> Reliability = number of agreement / (number of agreement + number of disagreement)

A high reliability score (0.87) was obtained for coding reliability of all data. Specific instances of disagreement were discussed until an agreement was reached, or both coders agreed to revise a code/label to resolve the disagreement.

Findings

This study identified a few key findings to answer the two research questions, 1) How much learning gains did students of diverse achievement levels, as represented by their different levels of prior knowledge of a designated topic, obtain in a structured peer assessment process? 2) What was the experience of students of diverse achievement levels in this peer assessment process?

Learning Gains

The first research question examined learning gains of students with diverse levels of content knowledge prior to peer assessment. This study attempted to answer this question from two perspectives: observed learning gains (as represented by point increase in grades between students' initial and final versions of WebQuest projects) and perceived learning gains (as represented by Likert Scale rating of task value in the post-assessment survey). As discussed previously, students were assigned to three groups: Beginning, Developing and Advanced, according to the grades of their initial submissions of WebQuest projects. Observed learning gains were determined by subtracting the points of their initial version

from the points of their final version. The average point gains in each group were calculated. Comparison of the grades (Figure 2) of student's initial and revised WebQuests in each group showed both the "Beginning" group and the "Developing" group gained the most significant portion of points (mean score of 9.5 and 7.05 points, respectively), while the "Advanced" group gained the minimum points (mean score of 1.75 points). In addition to the point gains, all 5 students originally in the "Beginning" level moved up to the "Developing" level. Eight students (out of ten) in the "Developing" level moved up to the "Advanced" level. Therefore, the number of students in the "Advanced" level increased from 6 to 14.

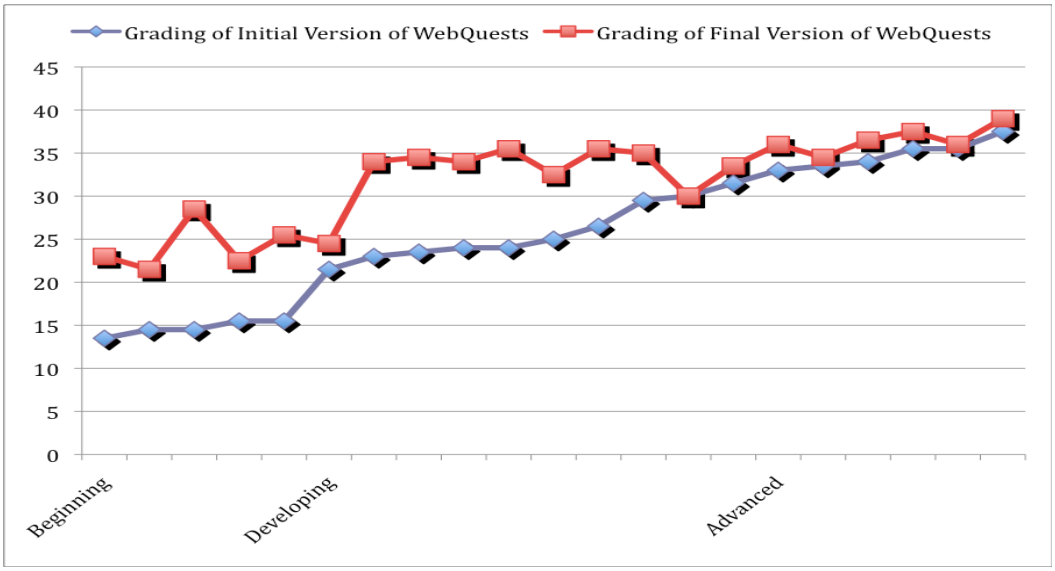

Figure 2. A Comparison of Student's Grades in Initial Version of WebQuests and Final Version of WebQuests

On the post-assessment survey, most students across the three levels (71%, n=15) stated they were satisfied with the quality of their own WebQuest projects, and most students (90%, n=19) provided positive or somewhat positive comments toward peer assessment. Nevertheless, when students were prompted to rate the quality of peer feedback they had received as Very Good, Good, Barely Acceptable, Poor or Very Poor, interestingly, most students in the "Beginning" (80%, n=4) and "Developing" levels (60%, n=6) rated good or very good, while only a small portion of students (16.7%, n=1) in the "Advanced" group rated the feedback they received as good. Furthermore, the researcher identified some interesting patterns from students' response to the seven 5-point Likert Scale statements regarding students' perceived learning gains.

Table 1. Student Perceived Learning Gains in Three Learning Groups (Beginning, Developing and Advanced) after Peer Assessment Activities

	Performance Level	Strongly Disagree	Disagree		Neutral		Agree		Strongly Agree	
			N	%	N	%	N	%	N	%
1). What I learned in this peer assessment activity is important.	Beginning						2	40%	3	60%
	Developing				3	30%	2	20%	5	50%
	Advanced						1	16.7%	5	83.3%
2). This peer assessment activity was useful in promoting my learning.	Beginning						2	40%	3	60%
	Developing						6	60%	4	40%
	Advanced						2	33.3%	4	66.7%
3). This peer assessment activity was interesting.	Beginning				1	20%	2	40%	2	40%
	Developing				1	10%	4	40%	5	50%
	Advanced		1	16.7%	2	33.3%	3	50%		
4). This peer assessment activity motivated me to learn.	Beginning						3	60%	2	40%
	Developing				4	40%	6	60%		
	Advanced				1	16.7%	2	33.3%	3	50%
5). It is valuable for me to be involved in this peer assessment activity.	Beginning				1	20%	1	20%	3	60%
	Developing				1	10%	4	40%	5	50%
	Advanced				1	16.7%	1	16.7%	4	66.7%
6). This peer assessment is a worthwhile activity.	Beginning						3	60%	2	40%
	Developing						6	60%	4	40%
	Advanced						2	33.3%	4	66.7%
7). I benefited from peers' comments.	Beginning				1	20%	1	20%	3	60%
	Developing				2	20%	4	40%	4	40%
	Advanced		1	16.7%	4	66.7%	1	16.7%		

Note: N=Number of Students; %=Percentage of Students; Beginning (n=5); Developing (n=10); Advanced (n=6)

View Table 1 JPG Full Size
<http://academics.georgiasouthern.edu/ijsotl/v5n2/articles/Li/images/Table01HiRes.jpg>

Table 1 summarizes students' responses to the task value statements in the post-assessment survey. As the table shows, most students in all of the three groups agreed that peer assessment was a worthwhile activity; peer assessment was useful in promoting their learning; and the peer assessment activity motivated them to learn. However, when asked if they felt peer assessment was interesting (Figure 2), 50% of students (n=3) in the "Advanced" group rated either "Disagree" or "Neutral", while most students in the "Beginning" and "Developing" groups (80%, n=4; 90%, n=9, respectively) held a positive attitude (rated "Agree" or "Disagree"). In addition, when rating the benefits of peers' comments (Figure 3), most students in the "Beginning" and "Developing" groups agreed or strongly agreed (both 80%) that they benefited from peers comments, while only 16.7% of students in the "Advanced" group said so. This finding was in line with students' rating of perceived quality of peer comments they received.

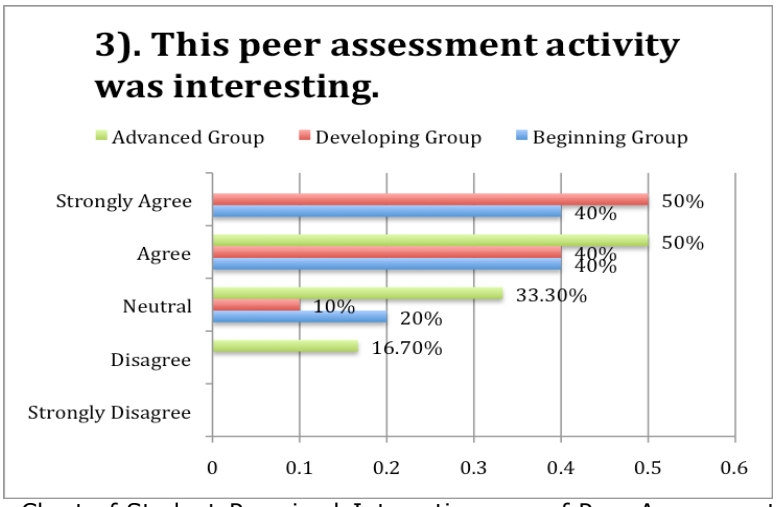

Figure 2. Bar Chart of Student Perceived Interestingness of Peer Assessment in Three Learning Groups (Beginning, Developing and Advanced)

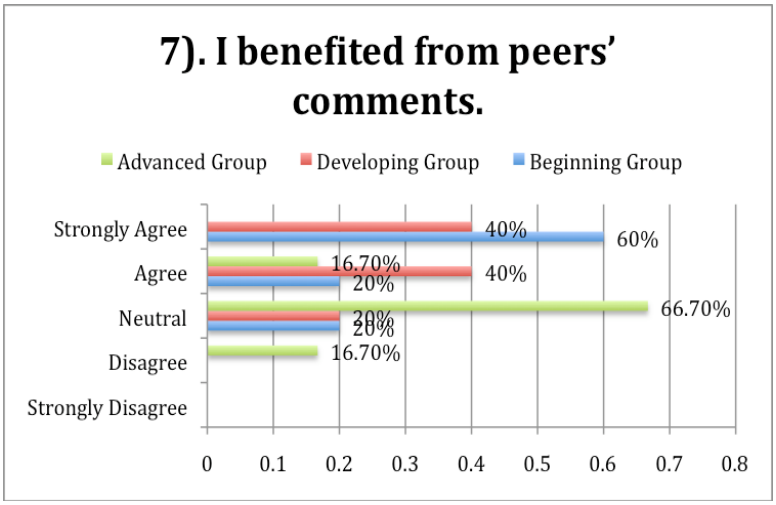

Figure 3. Bar Chart of Student Perceived Benefits of Peer Feedback in Three Learning Groups (Beginning, Developing and Advanced)

Two interpretations may be drawn from the picture of learning gains. One, student performance improved across the board; however, learning of high achieving students had not shown as much improvement as low achieving students. Two, low achieving students perceived more value from peer suggestions than high achieving students. These interesting findings indicated that learning gains (either observed or perceived) in peer assessment activities varied based on students' academic levels. On one hand, this may be explained by the "ceiling effect"—high achieving students had little room for performance improvement, as they already did a good job on their projects. One the other hand, this may suggest that high achieving students may not benefit from peer assessment activities as much as low achieving students.

Perceptions of Students in Different Learning Levels

Interviews were conducted to obtain more in-depth information regarding the students'

perceptions of peer assessment. The data analysis procedure resulted in the identification of three main themes. The first two (Quality of Peer Feedback Received and Confidence of Reviewing Peers' Projects) consist of students' positive, neutral or negative perceptions of peer assessment. The third theme (Impact of Peer Assessment Process) indicates mainly positive attitudes, although the focus varied by the achievement levels to which interviewees belonged.

Theme one—quality of peer feedback received.
Most interviewees expressed their satisfaction of the peer feedback received. They indicated they had received some constructive suggestions from peers and used them to improve their final WebQuest projects. Some direct quotes include:

> "I got some great suggestions from my peers. I believe that my work was largely improved…"
> "[Received] feedback helped me revise my WebQuest…"
> "The loops or the gap I had in my assignment I didn't see before until they brought them out, which I think my students would have found if I presented this to my own class, they too would have wanted more clarification. So, I thought they were very useful."

On the other hand, three interviewees expressed their disapproval or disappointment (more or less) of peer feedback they received. Beth (Beginning level) said that she got "overall good" comments. But some of them were a little "harsh." Aaron (Advanced level) said he liked the feedback received, although he did not agree with some of them—"I do not think this person understood my WebQuest." The other interviewee at the "Advanced" level, Allison, stated that she was disappointed with feedback and one set of feedback she received simply stated "Good job," which was meaningless.

> "One [set of] feedback was better than the other. One feedback kind of reminded me of the age of students I was targeting and made me kind of go back to simplify part of it… And the other feedback, it just said a good job. So I was kind of disappointed with that feedback."

It seems that interviewees in lower achievement levels had more positive attitudes towards the quality of peer feedback received than high achievers. The only complaint of peer feedback from Beth (Beginning level) was its lack of consideration, not the value; yet, both interviewed high achievers complained about the quality of peer feedback. This may not be a surprise as high achievers may have had a deeper and better understanding of what was required; therefore, they were able to critically judge peers' work and provide constructive suggestions for revisions. Low achievers, on the other hand, were probably still struggling with content. It may not be reasonable to expect them to pinpoint and analyze issues in peers' work in ways comparable to their high achieving classmates.

Theme two—confidence of reviewing peers' projects
Attitudes of interviewees varied drastically by whether they possessed critical assessment skills to assess peers' work. One interviewee in the "Beginning" level—Ben—did not think he was qualified to conduct peer assessment. He wrote that he had "never done this before", and he did not think he was ready to review peers' work. Two interviewees in the "Developing" and "Advanced" levels held positive attitude towards their assessment skills. They stated,

> "I guess I did all right. [When reviewing peers' projects], I saw a few problems that we had discussed in class …The assessment training and examples helped me a lot…" "Because we had gone over the expectations in class. And we did sample critiques of WebQuests in class. I kind of knew what was [a] good WebQuest as compared to [a] bad WebQuest…so I was confident and being able to provide valuable feedback to my peers."

The rest of the interviewees had somewhat neutral opinions about their assessment readiness. Although one interviewee deemed it "clumsy" for assessing peers' work, she felt okay to carry on these kinds of formative assessment activities, as long as her rating would not affect her peers' grades. Another interviewee said, "I don't know. [I] Guess I tried my best." Nevertheless, almost every one of the six interviewees, to some extent, accredited the training session they had before peer assessment for preparing them for assessment.

The interviewees' unanimous agreement towards the value of training confirms that training is a vital component of peer assessment. Researchers and educators have emphasized the importance of providing well-structured training to prepare students for critical peer assessment (Cheng & Warren, 1997; Dochy, et al. 1999). The major role for students in traditional education is as "learners". They are not ready to perform on the other side as assessors (Li, Liu, & Steckelberg, 2009). Therefore, training should be provided to maximize students' learning gains in peer assessment activities.

Theme three—impact of peer assessment process

Overall, students' appreciation of the peer assessment process was reflected across all of the interviews. Specifically, three interviewees in the "Beginning" and "Developing" groups commented on the positive impact of playing the different roles of assessors and assessees in peer assessment. Beth (Beginning group) indicated that, "I learned a lot from comparing my WebQuest with peers'…reading comments from peers about strengths and weaknesses of my work was also beneficial." Dana (Developing group) stated that, "I really think peer assessment should be used more in education. I learned a lot from this process…I enjoyed assessing others' work. Peer comments were timely, detailed and helpful. Sometimes I even feel that peer comments [may be more] helpful than teacher's feedback."

Interestingly, both the interviewees in the "Advanced" level emphasized the value more on the assessor aspect instead of the assessee aspect. Particularly, Alison said "When I see something that [is] unclear on peer's work, then I can go back to my own piece and make sure the objective there…" however, when she talked about peer comments, she stated, "I wasn't incredibly pleased with the result or the feedback I got. So I think it could be better feedback. But the one good response I did get did make me go out and take out some of the overly specific details…"

It is quite interesting to interpret the meaning of the data and define the message they want to convey. As discussed previously, analysis of students' response to the post-assessment survey suggested that high achievers perceived less value in peer feedback as compared to low achieving students. These interview quotes further echoed that finding. The triangulation of data analysis provides a picture of how achievement levels affect students' perceptions of peer feedback value.

Discussion

The purpose of this study is twofold: (1) to determine how much peer assessment may benefit students of diverse achievement levels and (2) to explore perceptions of students, of diverse achievement levels, in regard to this peer assessment process. Our findings indicated that students in the "Beginning" and "Developing" levels benefited most (in terms of points gained) from the peer assessment process. Students in the "Advanced" category had the least points gain. Survey responses suggested that all students, regardless of their initial performance levels, recognized the value of peer assessment and generally agreed that peer assessment was a worthwhile learning experience. Nevertheless, when rating the interestingness of the peer assessment activity and the value of the peer comments received, students in the "Advanced" group were less satisfied as compared to those in the other two groups. This picture was echoed later by the interviews with selected students from each of the three groups.

The findings of this study have testified the previously stated speculation that peer assessment may have a different impact on students of diverse achievement levels (Li & Steckelberg, 2005; Davies, 2001), and have implied that students in early developmental stages may benefit more from the peer assessment process than high achieving students in terms of score increases. These findings are in line with previous research reporting that students of various ability levels had various learning gains in collaborative learning settings (Webb, 1989, 1991; Webb, Nemer, Chizhik & Sugrue, 1997). For example, Webb, Nemer, Chizhik and Sugrue (1997) examined the effects of group ability composition on group processes and student learning outcomes. Their findings suggested that below-average students, when working in groups with above-average students, had performed better on achievement tests as compared to below-average students working in groups without above-average students. Roberts (2004), in the Preface of the book "Online collaborative learning: theory and practice", summarizes that "research has suggested that group work tends to advantage below-average students" (P. vii).

If this implication is true, why did high-achieving students still feel positive about peer assessment? If their learning gain in peer assessment was not significant, why did they still rate peer assessment as a "worthwhile" activity? Researchers suggest that peer assessment activities involve multiple learning stages, and each of these stages may benefit students (Li, Liu & Zhou, in press). Is it possible that, although high achieving students may not receive much constructive peer feedback as assessees, serving as assessors and helping beginners may deepen high achievers' understanding of the content area and strengthen their reflection. As Topping (1998) posited, assessment may benefit both assessors and assessees in multiple domains such as "cognition and metacognition, affect, social and transferable skills, and systemic benefits" (p. 254). Future studies are warranted to further examine this intriguing observance.

Students' survey responses and interviews suggested the most compelling difference between the 'Beginning" and "Developing" groups, and the "Advanced" group was the students' different attitudes towards value of peer feedback they received. Most students in early developmental stages embraced peer comments received, while adept students showed disappointment regarding the quantity and quality of peer feedback. As Allison in the "Advanced" group said during her interview, one of her peers simply gave her "Good job" commentaries, which she deemed "meaningless". Another interviewee in the same group, Aaron, stated that he did not think one of his peer reviewers understood his WebQuest.

To better understand this scenario, the researcher went back to review Allison and Aaron's initial and final versions of their WebQuest projects and the peer comments they received. Allison, as one of the most competent students in her class, was given 38 (out of 40) for her initial submission and awarded the full mark for her final submission. Not surprisingly, her peers did not find much to suggest for further improvement. It is also important to note that one of Allison's peers did identify an issue on her initial WebQuest, which, upon her correction, led to the two-point increase on her final submission.

Aaron's story was different. Two students who were in the "Beginning" and "Developing" groups reviewed his project. One of the reviewers apparently was not clear about the critical elements of WebQuests and provided some misleading suggestions. This explained why Aaron complained that she did not understand his project. While these complaints from students (especially high achieving students) are probably inevitable in peer assessment, it is critical to design peer assessment activities in ways that minimize the negative impact of these issues. For example, students should be educated that peer feedback they receive may vary in quality. It is a learning process for them to critically judge the quality of peer feedback they receive and make decisions to adopt or not adopt a suggestion. In addition, students should also learn that receiving peer feedback is neither the only step nor the ultimate step in peer assessment activities. Instead, peer assessment may benefit students (either as assessors or assesses) in multiple ways, such as increased time on learning task, focused attention on critical elements of quality work, and greater sense of accountability and responsibility (Topping, 1998). Students will achieve maximum gains from peer assessment only after they fully understand this process and its potential.

It is worth noting that training played a vital role in preparing students for peer assessment in this study. Even though more than half of the interviewees (n=3) still held either negative or neutral thoughts regarding their readiness for peer assessment, nearly all of them favorably commented about the importance of the pre-assessment training in regard to preparing them for the assessment task. This further confirmed that training is a critical component of peer assessment, as claimed by various researchers (Chen & Warren, 1997; Dochy, et al., 1999; Van Zundert, et al., 2010). Future studies should explore more ways to effectively train and prepare students before engaging them in critical peer assessment.

Based on literature and the findings of the current study, the researcher recommends that training should be provided in future peer assessment studies to prepare students and to maximize students' learning gains. Students need to understand the significance of peer assessment as well as acquire critical assessment skills before they can be engaged in any kind of assessment activities. In the implementation process of peer assessment, the researcher also suggests that students' learning levels should be carefully considered when placing students into groups. It may not be an ideal composition to have groups of all low or all high achievers. Future studies should look into the relationship between student learning gains, perceptions and group organizations.

Conclusion

This study is important in its nature in that it is one of the first few to examine how peer assessment may benefit students of various achievement levels. Although the value of peer assessment has been studied and documented by a growing body of research across the world, it is still unclear how this strategy works and what constitutes effective peer assessment (Van Zundert, et al., 2010). This study utilized a mixed methodology approach, seeking to get a clear picture of the "how students benefit" issue from various perspectives. Interview comments were integrated with the quantitative data of students' project quality

and the post-assessment survey, thereby providing a vivid portrait of learning gains and perceptions of students at varied initial academic levels.

There were a few limitations associated with this study. One limitation was that purposeful sampling was used and participants were selected because they were graduate students enrolled in a technology application course. Generalizability should not be assumed in other settings and using other populations. Future studies should recruit a different sample to verify results. Another limitation was related to the small sample pool. This study involved 21 students who were classified into three groups based on their initial performance on their WebQuest projects. Group sizes were small, ranging from 5 to 10. Due to this limitation, the researcher was only able to compare mean differences of learning gains (both observed and perceived) among three groups at a descriptive level. No inferential statistical analysis was conducted to test the significance of the observed differences. Future studies should attempt to increase the sample size to investigate the extent to which students' learning gains vary among the three groups.

References

Bellofatto, L., Bohl, N., Casey, M., Krill, M., & Dodge, B. (2001, June 19). *A rubric for evaluating WebQuest*. Retrieved from http://webquest.sdsu.edu/webquestrubric.html

Brown, D. (2004). *Language assessment: Principles and classroom practice*. New York: Longman.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-238.

Cheng , W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of group project. *Assessment & Evaluation in Higher Education*, 24(3), 301-314.

Chenitz, W. C., & Swanson, J. M. (1986). *From practice to grounded theory: Qualitative research in nursing*. Menlo Park, CA: Addison-Wesley.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International*, 37(4), 346–354.

Dochy, F., Segers, M., & Sluijsmans, D.M.A. (1999). The use of self-, peer-, and co assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350.

Dodge, B. (1995). WebQuests: A technique for internet-based learning. *Distance Educator*, 1 (2), 10-13.

Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32(2), 175-187.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London, Routledge Falmer.

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In Brown, S. & Glasner, A. (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp. 41-53). Buckingham: Open University Press.

Hiltz, S. R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, 40(2), 44-49.

Johnson, R.B., & Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33 (7), 14-26.

Li, L., & Steckelberg, A. L. (2004). Using peer feedback to enhance student meaningful learning. *Proceedings of the Association for Educational Communications and Technology*, Chicago, IL. Accession No: ED485111.

Li, L. (2007). Effects of critical assessment training on quality of peer feedback and quality of students' final projects in peer assessment (Doctoral dissertation). Available from *Dissertations Abstracts*. (AAT No. 3271906).

Li, L., & Steckelberg, A. L. (2005). Impact of technology-mediated peer assessment on student project quality. In Simonson, M. & Crawford, M. (Eds.), *Proceedings of the Association for Educational Communications and Technoloy International Conference 2005* (pp. 307-313). Orlando, FL.

Li, L., Liu, X., & Steckelberg, A. L. (2009). Analyzing peer feedback in a technology-facilitated peer assessment. In C. D. Maddux (Ed.), *Research Highlights in Technology and Teacher Education 2009* (pp. 213-222). Chesapeake, VA: AACE.

Li, L., Liu, X., & Zhou, Y. (In Press). Give and take: A re-analysis of assessor and assessee roles in technology-facilitated peer assessment. *British Journal of Educational Research.*

Li, L., Steckelberg, A. L., & Srinivasan, S (2009). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology*, 34 (2), 133-148.

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2002). Student attitudes toward networked peer assessment: Case studies of undergraduate students and senior high school students. *International Journal of Instructional Media*, 29(2), 241-254.

MacGregor, S., & Lou, Y. (2004). Web-based learning: How task scaffolding and web site design support knowledge acquisition. *Journal of Research on Technology in Education*, *37*(2), 161-175.

March, T. (2011, June 13). *Criteria for assessing best WebQuests*. Retrieved from http://bestwebquests.com/bwq/matrix.asp

Martin, B. (1999, August 28). *WebQuest evaluation rubric*. Retrieved from http://drwilliampmartin.tripod.com/s_WebQuest_Rubric1.html

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: SAGE

Patri, M. (2002). The influence of peer feedback on self and peer-assessment of oral skills. *Language Testing*, 19(2), 109-131.

Pope, N. (2001). An examination of the use of peer rating for formative assessment in the

context of the theory of consumption values. *Assessment & Evaluation in Higher Education*, 26(3), 235-246.

Roberts, T.S. (2004). *Online collaborative learning: Theory and Practice*. Hershey, PA: Information Science Publishing.

Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18, 221–233.

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

Topping, K., Smith, F. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, 25(2), 149−169.

Van Zundert, M., Sluijsmans, D. M. A., & Van Merrie¨nboer, J. J. G. (2010). Effective peer assessment processes: research findings and future directions. *Learning and Instruction,* 20(4), 270-279.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21-39.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal of Research in Mathematics Education*, 22, 366-389.

Webb, N. M., Nemer, K., Chizhik, A., & Sugrue, B. (1997). *Equity issues in collaborative group assessment: Group composition and performance*. Center for the Study of Evaluation, Graduate School of Education, UCLA, Report to the Office of Educational Research and Improvement/ Department of Education. Retrieved April 16, 2011 from http://www.cse.ucla.edu/products/Reports/TECH457.pdf

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(3–4), 186–193.