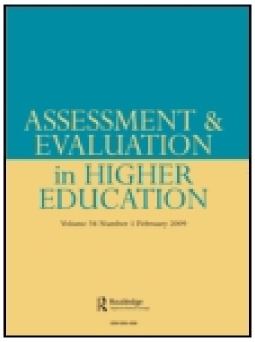

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: https://www.tandfonline.com/loi/caeh20

The effect of peer assessment on project performance of students at different learning levels

Lan Li & Fei Gao

Assessment & Evaluation in Higher Education, 2016
Vol. 41, No. 6, 885–900, http://dx.doi.org/10.1080/02602938.2015.1048185

The effect of peer assessment on project performance of students at different learning levels

Lan Li[a]* and Fei Gao[b]

[a]*School of Teaching and Learning, Bowling Green State University, Bowling Green, OH, USA;* [b]*Department of Visual Communication and Technology Education, Bowling Green State University, Bowling Green, OH, USA*

Peer assessment has been increasingly integrated in educational settings as a strategy to foster student learning. Yet little has been studied about how students at different learning levels may benefit from peer assessment. This study examined how peer-assessment and students' learning levels influenced students' project performance using a two-way factorial design. One hundred and thirty teacher education students participated in this quasi-experimental study. When working on a technology-integrated lesson plan project, the experimental group completed an online peer assessment process while the control group followed the discussion method. Students' learning levels were measured and divided into low, average and high achieving according to the quality of their draft lesson plans. Data analysis suggested that the impact of peer assessment on students' lesson plan project seemed to vary according to students' learning levels. While low- and average-achieving students showed significantly improved performance right after the integration of a peer assessment model, the model seemed to have had less impact on the performance of high-achieving students. Significance, implications and limitations of findings are discussed.

Keywords: peer assessment; students learning levels; performance; feedback

Overview of peer assessment

Peer assessment is the process of students evaluating each other's work using performance criteria (Falchikov 2007). Although existing in many variants, peer assessment usually involves students reviewing each other's work for formative feedback, summative grading or a combination of both. As an effective learning tool, formative peer assessments are often utilised in curricula by engaging students in both roles as 'assessor' and 'assessee.' As assessors, students review peers' work and provide feedback; as assessees, students read and act upon peer feedback received to improve their own work (Li, Liu, and Steckelberg 2010; Li, Liu, and Zhou 2012).

The value of peer assessment exists, in part, in its ability to engage students in the learning process and encourage self-assessment and reflection. One of the main goals of education is to prepare students as adaptive and reflective practitioners who are critical thinkers and problem solvers (Falchikov and Boud 1989; Kwan and Leung 1996). Being able to make informed judgements about their own and peers' strengths and weaknesses is an important skill that will help students succeed in both education and the workplace (McDonald et al. 1995). Peer assessment is seen

*Corresponding author. Email: lli@bgsu.edu

as a means by which this vital lifelong learning skill is acquired and developed (e.g. Brown, Rust, and Gibbs 1994).

Peer assessment has been widely applied across diverse academic contexts, such as teacher education (Li 2012), computer science (Wang et al. 2012), medicine (Violato and Lockyer 2006), engineering (Hersam, Luna, and Light 2004), second-language learning (Cheng and Warren 2005), biology (Orsmond, Merry, and Reiling 1996) and business (Brutus, Donia, and Ronen 2013). Numerous studies have demonstrated positive effects of peer assessment on student academic achievement, cognitive growth and other outcomes. Some potential learning and developmental benefits reported include increasing students' autonomy and motivation (Brown 2004; Hiltz and Wellman 1997; Pope 2001) and responsibility (Somervell 1993), promoting students' critical thinking skills, encouraging interpersonal skills among students, deepening students' understanding of assessment criteria and quality performance (Dochy, Segers, and Sluijsmans 1999; Patri 2002), providing students with adequate and timely feedback (Gibbs 1999) and enhancing learning outcomes (Li and Steckelberg 2005; Pope 2001). After an extensive review of 109 peer assessment studies conducted in higher education, Topping (1998) concluded that, when developed and implemented appropriately, peer assessment can 'yield gains in the cognitive, social, affective, transferable skill, and systemic domains that are at least as good as those from staff assessment' (269).

As with any instructional strategy, the application of peer assessment is not without challenges. One of the drawbacks is peer pressure. Some students feel uneasy and reluctant to assess their peers' work and 'marking could be easily affected by friendship, cheating, ego or low self-esteem' (Robinson 1999, 96). Another possible pitfall is the demand of time. Peer assessment is a 'daunting' task and requires effort and commitment from both instructors and participating students. From the perspective of instructors, managing a peer assessment process requires substantial time, especially when confidentiality of assessors and assessees is required. From the perspective of students, additional time on tasks such as attending assessment training, reviewing peers' work, and judging the value of peer feedback is warranted to complete a peer assessment activity.

In addition to peer pressure and demand on time, Li, Steckelberg, and Srinivasan (2008) also highlight two more barriers: students' ineptitude in understanding assessment criteria and their incapability of undertaking assessment. Students can conduct valid assessment and provide valuable feedback only when they understand how quality performance is defined. However, students may lack the ability of correctly interpreting and applying marking criteria, which would impact the validity and reliability of peer marking. Apart from the problem of measuring quality performance, students may also be inexperienced and ill equipped to identify the strengths and weaknesses of peers' work. All of the aforementioned downsides may compromise the effects of peer assessment. These are daunting yet not insurmountable challenges and require some thoughtful and intentional consideration. Therefore, implementing peer assessment in a curriculum is not a decision that should be taken lightly; and there is a need to address and resolve related issues before adoption.

Technology-enhanced peer assessment

In the past two decades, the advancement of technology has created unprecedented possibilities for peer assessment. Software tools that allow students to actively participate in the process of peer assessment not only provide promising solutions for the challenges paper-based peer assessment face, but also create new features and dimensions not available in conventional systems. One of the primary advantages that technology-enhanced peer assessment can offer is anonymity (Li, Steckelberg, and Srinivasan 2008). Conducted in an anonymous system, this method allows for identities of assessors and assessees to be easily disguised by user names or pseudonyms. When students do not know whose work they are marking or who feedback is from, peer pressure is greatly reduced, thus increasing the reliability and validity of peer marking (Zhao 1998). Another improvement that technology-enhanced peer assessment can make is ease of management. Peer assessment can be time consuming to administer in a paper-based system. An earlier study (Hanrahan and Isaacs 2001) reported more than 40 person hours for managing an anonymous peer assessment distribution system with 244 students. In a technological environment, since data are managed by means of an integrated database system, there is limited manual work needed to maintain the assessment distribution system, such as assigning students' work to review, collecting peer feedback from assessors and distributing feedback to assessees. The management workload for technology-integrated peer assessment is minimum.

Besides these benefits, technology can also add values in other aspects such as convenience of access and accommodation of multimedia files. Online tools can support peer assessment in an anytime and anywhere manner (Chen 2010). Reviewing peer work, and providing and receiving feedback are no longer confined to the four walls of a classroom, which allows students to engage in active learning at their own pace and also provides teachers with flexible opportunities to manage and monitor peer assessment systems at their convenience. Technology is also noted for its capability of accommodating multimedia files such as audio recordings and videos (Liu and Li 2013; Sung et al. 2005). The format of assessed work and peer feedback is no longer restricted to text, but can incorporate various elements such as pictures, animations, videos, audio recordings and interactive simulations. The coexistence of these versatile learning components attends to individual needs and preferences and can be easily managed in computerised systems. To sum up, compared to paper-based systems, technology-enhanced peer assessment has considerable potential.

Underlying theories

Although it is difficult to identify a theory underlying peer assessment due to its great variations, Topping (1998) posits that peer assessment might be grounded on social constructivism (Vygotsky 1978). The core tenet of social constructivism is the belief that knowledge is advanced through learners' meaningful interactions in social contexts and cultural settings (Bauman 2012). Vygotsky (1978) suggests that learning has social origins and an individual's development cannot be understood without considering its social and cultural contexts. He asserts that higher mental functions occur when learners interact with each other and the outside world (Vygotsky 1978).

One of the key concepts of Vygotsky's social constructivism theories is the More Knowledgeable Other. Vygotsky (1978) defined it as someone who has higher skills or more experiences than the learner with regard to a certain task. Teachers or older adults like parents are usually the ones with more knowledge or experiences. However, the more knowledgeable other can also be a peer, or even a computer programme like an electronic tutor (Maclellan and Soden 2008). In a scaffolding instruction, a more knowledgeable other provides support to facilitate a learner's development (Van Der Stuyf 2002).

Another key concept of social constructivism that goes hand in hand with the More Knowledgeable Other is the Zone of Proximal Development. Vygotsky (1978) defined the 'zone' as the distance between the 'actual development level' of a learner and the higher level of 'potential development' attainable through 'adult guidance or in collaboration of more capable peers' (86). In other words, the zone of proximal development speculates that learners build and test their knowledge through dialogues with more competent and experienced others. From the perspective of the Zone of Proximal Development, the main goal of education, therefore, was to keep learners in their own zones as much as possible in order to achieve the maximum learning gain (Espinoza and Winsler 2005).

The social constructivist approach, including the More Knowledgeable Other and the Zone of Proximal Development concepts, is involved in most formative peer assessment models where students first act on what they can do independently; and then with assistance from peers, teachers or other supporting systems, advance their concept knowledge and improve the quality of their work. The interplay between students during peer assessment promotes learning and skill acquisition; and less able students are able to raise their competence through help from more able peers.

As social constructivism believes that learning is the dynamic interplay between individuals and the environments they are situated in, recent research on peer assessment suggests that the interaction with more capable peers in peer assessment may not be the sole determinant of benefits. Learning gains may also occur when students are engaged in peer assessment activities such as defining marking criteria, providing feedback to others or viewing other students' projects. Collectively defining and clarifying an assessment rubric in cooperation with teachers may increase students' awareness of their own performance (Freeman 1995; Mehrens, Popham, and Ryan 1998). Studies (e.g. Li, Liu, and Steckelberg 2010) also suggest a significant relationship between the quality of feedback students provide to peers and the quality of their own projects. Further, reviewing peers' performance may lead to observational learning, also known as social learning, which is defined by Bandura (1977) as learning that occurs through observing behaviour of others.

Impact of peer assessment and student learning levels

Research on peer assessment during the last a few decades has consistently documented the positive effects of peer assessment, especially its impact on learning. Yet few reported studies evaluated the impact of peer assessment on students with various achievement levels. In other words, would peer assessment have equal impact on all students? How may peer assessment benefit students of diverse achievement levels? Or would low-achieving (LA), average-achieving (AA) or high-achieving (HA) students respond differently to peer assessment? It remains unclear what makes effective peer assessment (Van Zundert, Sluijsmans, and Van Merriënboer 2010).

One of the few studies that draw attention to the 'how students may benefit' perspective of peer assessment was conducted by Davies (2000). While the main goal of Davies's study was to report the functionalities of a computerised peer assessment system and its effectiveness from students' perceptions, he observed that the peer-assessment process had a different impact on students of diverse achievement levels. Specifically, comparison of pre- and post-test scores suggested that HA students may have benefited least (gained the fewest points on the post-test, compared to scores of their pre-test) from their peer assessment activities, while LA students statistically benefited most (gained the most points on the post-test, compared to scores of their pre-test).

A different study suggesting a possible unequal impact of peer assessment on student learning was conducted by Li and Steckelberg (2004), who randomly assigned students into two groups (experimental and control) after they completed the initial draft of their project. While students in the experimental group reviewed and rated peers' projects to help improve each other's projects, students in the control group worked on improving their projects. The evaluation of students' initial project and revised project suggested that the final scores of the control group had a much smaller variability compared to those of the control group. Based on this finding, the researchers posited that the learning gains of peer assessment may vary for students at different learning levels. This postulation was supported in one of their later studies (Li, Steckelberg, and Srinivasan 2008), which explored student perceptions towards peer assessment. Students' survey responses noted that one HA student commented upon the impact of peer assessment, 'Sometimes peer assessment isn't helpful if you already did a good job' (143).

These interesting results have drawn the researchers' attention to an important aspect of peer assessment: how this approach may benefit individual students at different learning levels. Surprisingly, a literature search indicated that this characteristic of peer assessment has not yet been adequately studied.

Li (2012) took the argument a step further and examined how students of different achievement levels may benefit from peer assessment with a mixed methodology approach. After 21 students completed the initial draft of their projects, they were categorised into three groups (LA, AA and HA) based on the quality of their draft projects. All students went through the same peer assessment process, reviewed and commented upon each other's projects, and then viewed feedback and improved their own projects. Findings supports previous studies in that students in the early academic development phases (LA and AA) benefited more from peer assessment than students in the advanced level in terms of points gained in their revised projects after peer assessment. Interestingly, the students' survey responses and follow-up interviews suggested that students across the board, despite their academic levels, generally recognised the value of peer assessment and rated it as a worthwhile activity. The researcher suggested that this may be due to students' deeper and better understanding of peer assessment as a reflective learning tool instead of just a step to acquire peer feedback. Intriguing as these findings were, the small sample size utilised in the study (21 students) did not provide enough statistical power to conduct inferential analysis. The significance of this study is limited since mean differences of student learning gains among achievement groups could only be compared at a descriptive level.

Research questions

Drawing upon the findings and indications of previous studies, this study aimed to investigate effects of peer assessment and student achievement levels on students' project performance in an online peer assessment model. The researchers were especially interested in examining the possible interplay between peer assessment and students' learning levels on students' performance on a lesson plan project. As students learn in different ways and at different paces, assessment of student learning takes various forms. For example, assessment can be formative or summative, qualitative or quantitative, direct or indirect, standards-based or value added: thus, in his toolkit for lectures, Race (2007) examined 15 different approaches of assessing learning. The concern of this study only focused on students' immediate learning outcome (student project performance measured by project scores) right after peer assessment. Other types of learning were not addressed.

The specific research questions of the study were the following:

(1) Are students who complete an online peer assessment more likely to perform significantly better on a technology-integrated lesson plan project than those who do not?
(2) Do peer assessment and students' learning levels interact in their influence on student performance on a technology-integrated lesson plan project?
(3) Are LA, AA and HA students who complete an online peer assessment more likely to perform significantly better on a technology-integrated lesson plan project than their counterparts who do not complete the online peer assessment?

Methodology

Participants, study context and design

Participants in this quasi-experimental study were 130 undergraduate teacher education students enrolled in a technology application course at a major midwestern public university in the United States. All participants were traditional students in a highly competitive, four-and-a-half year educational programme. Admission to the programme requires completion of 45 credit hours and a minimum Grade Point Average of 2.8 (on a 4.0 scale). The purpose of the technology course was to help teacher candidates develop technology skills and classroom technology integration strategies. The face-to-face class lasted 16 weeks and used blackboard, which is a course management system that can be used to display course materials, and enables both instructors and students to interact with each other and with course materials online.

This study employed a between-subjects 2 × 3 factorial design with peer assessment (with or without peer assessment) and student achievement levels (LA, AA and HA). The dependent variable was students' performance on a technology-integrated lesson plan project, which demonstrated students' capability to design technology into curriculum to enhance instruction. Students were assigned into either the peer assessment group ($n = 63$) or the discussion board group ($n = 67$) based on the sections they were enrolled in. Additionally, based on scores of their draft lesson plan (maximum points = 40), students were allocated into one of the three learning levels: LA, AA and HA. Students who scored 30 and higher were assigned into the HA group ($n = 43$); students who scored between 20 and 30 were grouped into the AA group ($n = 45$); the rest of the students were assigned into the LA group ($n = 42$).

Self and peer assessment building block

The Self and Peer Assessment Building Block (SPABB) in Blackboard was utilised to facilitate peer assessment in this study. With SPABB, creating a self-assessment or peer assessment, or a combination of both, is fairly straightforward. With the course management system, instructors can easily create or import an assessment and control various assessment parameters, such as date ranges (e.g. how long an assessment will be available to students, when students can submit their work for assessment, and when students should complete and submit their assigned assessment), anonymity, number of submissions to assess, etc. With SPABB, instructors first set up the peer assessment task. Students can then submit their work for assessment within the date range for submission. Once the assessment starts, students' submitted work is randomly assigned to a number of peers for review as determined by the instructor. After the assessment activity completes, students are able to view peer feedback provided to their own work in their own accounts. The whole assessment system is managed online within the course management system. Administration workload is minimum and anonymity of assessors and assessees is ensured.

Instructors have direct control over the peer assessment process. During each of the submission and peer assessment phases, instructors are able to monitor students' submissions as well as the peer assessment process. Submissions of students work as well as the assessment/reviews students provided to their peers could also be downloaded as individual files for later use.

Procedure

In the two-week-long lesson plan module, all students completed a technology-integrated lesson plan project, which required students to design instructional activities that integrated technology into their specific content areas and grade levels. Students were first instructed to identify a target school district through an Internet search and then investigate the availability of technology in their target school districts. Their lesson plans should address the state and national content and technology standards, curriculum guides of their target school districts and other resource requirements. Their lesson plans should also demonstrate appropriate use of technology and its impact on curriculum, instruction and assessment. After students completed their draft lesson plans, the peer assessment group followed the following steps:

Step 1: Teacher explicated aims and expectations

Students first watched a short video introducing peer assessment and summarising its potential learning benefits and implementation challenges. The teacher then reinforced the idea that the purpose of the peer assessment activity was to foster learning, and encouraged students to provide 'true' evaluation of peers' performance. Those students who were reluctant to rate peers or afraid of receiving 'unfair' markings from peers were assured that peer rating would only be used for the purpose of project improvement and would not be used to adjust students' final project scores. Instead, students' evaluation process (how accurately they marked peers' work) would be evaluated by the teacher and contribute to their own final project scores. Thus, intentionally over-marking or under-marking could potentially lower students' own scores.

Step 2: Students defined marking criteria with each other and the teacher

The teacher first provided students with the project rubric and then instructed students to review and discuss it. Afterwards, students practiced assessing two example lesson plans using the rubric. Their ratings and the teacher's rating were compared and discussed to enhance students' understanding of the project and the marking rubric.

Step 3: Students created projects

Students developed the lesson plan in a Word document based on the rubric.

Step 4: Students analysed peer work and provided feedback

This study utilised SPABB in blackboard to allow students to review and grade peers' lesson plans online. Students uploaded their completed lesson plan documents to the blackboard course shell. After the assignment submission was complete, the peer review process started. Students were randomly assigned to review two other students' lesson plans. In this stage, each student played two roles: assessor and assessee. As assessors, students read assigned lesson plans, allocated points and provided qualitative feedback to justify their ratings or suggest improvements. As assessees, students received points and corresponding feedback from two peers with regard to their own lesson plans. During this stage, confidentiality was enforced, and the identifications of assessors and assessees were concealed.

Step 5: Students assessed the value of feedback received

As assessees, each student received scores and qualitative feedback from two peer assessors regarding their own project. Students were told clearly that not all peer reviews have equal value and feedback may vary in quality (Li 2012). Students were reminded that, although they should consider all feedback, they should not suspend their own judgement and blindly follow feedback received.

Step 6: Students enhanced their own work

After students assessed the value of feedback received, they were now in a position to decide how to respond. Students were instructed to only follow suggestions that they deemed valid, appropriate and helpful.

Students in the no peer assessment group followed the conventional method to learn and interact. Students participated in a class discussion regarding the general project requirements, as well as their group discussions to address specific topics concerning their content areas or grades. They were encouraged to communicate with the instructor for any assistance they needed, and opportunities for optional individual meetings with the teacher were also provided.

Lesson plan rubric

The lesson plan rubric included eleven criteria and was organised into five sections: standards and objectives, materials and technology, procedures, assessment and differentiation. Each criterion of the rubric included three performance indicators (unacceptable, acceptable and target) and corresponding points. The maximum possible points a lesson plan could receive was 40.

Grading

To ensure grading reliability, one independent rater collaborated with the instructor to grade all lesson plans. They first participated in an eight-hour training, which included two rounds. In the first round, they first discussed the rubric to reach a general consensus of the measuring criteria. They then applied the rubric to evaluate three example lesson plans: one target, one acceptable and one unacceptable. Afterwards, they compared and discussed the variations of their grading of these three examples until all disagreement were resolved. In the second round, three more example projects in each performance level were provided for the instructor and independent rater to practice grading. The instructor and the rater compared their evaluations and discussed the variation of their evaluations until all disagreements were resolved. Upon the completion of grading training, both the instructor and the grader rated all lesson plans. Pearson's correlation (r) between scores given by the instructor and the independent grader was highly satisfactory, 0.87, which indicated the grading of lesson plans was consistent and reliable. The mean scores of the instructor and the rater's grading were used in data analysis.

Results

This study used a quasi-experimental design and students were assigned into the peer assessment and no peer assessment groups based on the class sessions they were enrolled in. The two groups were equivalent in terms of students' scores of draft lesson plans, $t(128) = .893$, $p = .373$. Levene's test for homogeneity of group variance for the scores of students' lesson plans was significant ($p = .002$). To address such heterogeneity, the researchers adjusted the p value by assuming a more stringent significance level 0.01 instead of the commonly used threshold value of 0.05, as suggested by Keppel and Wickens (2004), to decrease the possibility of declaring statistical significance by chance.

Main effect of peer assessment & interaction between peer assessment and student learning levels

Table 1 demonstrates the mean scores and standard deviations of student performance (in terms of project scores) on their lesson plans. Table 1 shows that student

Table 1. Mean scores and standard deviations of student scores on lesson plan by groups and student learning levels.

Groups	No peer assessment			Peer assessment		
Learning levels	LA	AA	HA	LA	AA	HA
Lesson plan score						
M	21.43	28.05	34.48	27.19	31.38	36.06
SD	3.40	2.58	1.49	4.38	3.73	1.98

Note: LA – Low-achieving; AA – average-achieving; HA – high-achieving; *M* – mean; SD – standard deviation.

Table 2. Experimental group (peer assessment) × learning levels factorial ANOVA for lesson plan.

Source	Df	F	η^2	p
(A) Peer assessment	1	42.597	.256	0.000
(B) Learning levels	2	132.046	.680	0.000
A × B (interaction)	2	4.863	.073	0.009
Error (within groups)	124			

performance on their lesson plans in the peer assessment group, in numerical values, is higher than that of the no peer assessment group across all levels.

A 2 × 3 factorial analysis of variance (ANOVA) was conducted to evaluate the effects of peer assessment and students' learning levels on student performance on their lesson plans. As Table 2 shows, the interaction between levels of student learning and peer assessment was significant ($F = 4.863$, $p < .01$, $\eta^2 = .073$), which indicated that at least one interaction effect existed. The result suggested that the magnitude of the difference between the means of the peer assessment and no peer assessment groups depended upon student learning levels. Simple main effects analysis for peer assessment was significant ($F = 42.597$, $p < .001$, $\eta^2 = .256$), which suggested that students who were in the peer-assessment group outperformed students in the no peer assessment group on the lesson plan project, ignoring the effect of student learning levels.

Comparison of project performance of LA, AA and HA students who complete peer assessment with their counterparts who do not

To obtain more focused, specific information on where differences are in the interaction effect, one-way ANOVA and a cell-means model were used to compare the differences of the six treatment combinations, which were coded as 1 (peer assessment + LA), 2 (peer assessment + AA), 3 (peer assessment + HA), 4 (no peer assessment + LA), 5 (no peer assessment + AA) and 6 (no peer assessment + HA).

To examine the difference among the six group means, one-way ANOVA analysis was conducted. The results indicated at least one group mean was different from the others, F (5, 124) = 62.741, $p < .001$. Since equal variances and sample sizes were not assumed, Game–Howell was used for *post hoc* comparisons (ANOVA Options n.d.). As Tables 1 and 3 show, first, students of LA level in the peer assessment group ($M_1 = 27.19$) performed significantly better on their lesson plans than their counterpart ($M_4 = 21.43$) in the no peer assessment group (1 vs. 4 in Table 3, $p < .001$). Second, according to the adjusted .01 p value due to unequal variances, data analysis also suggested a statistically marginal difference ($p = .013$) between the AA students in the peer assessment group and no peer assessment group (2 vs. 5 in Table 3). That the group mean of Group 5 was higher than that of Group 2 (M_5: 31.38 > M_2: 28.05) indicated that AA students who follow the web-based peer-assessment model performed significantly better on the lesson plan project than their counterparts who did not. Third, nevertheless, analysis of HA students' lesson plans of both groups showed no statistically difference (3 vs. 6 in Table 3).

Table 3. Game–Howell pairwise comparisons of cell means of lesson plan project of three learning levels in peer assessment group and no peer assessment group.

Mean	VS	Mean	Sig.	Mean difference	Standard error of difference	Lower bound	Upper bound
1	VS.	4	.000	−5.762	1.209	−10.16	−1.37
2	VS.	5	.013	−3.327	.946	−6.74	.09
3	VS.	6	.072	−1.576	.547	−3.61	.46

Notes: 1 = No peer assessment + low achieving; 2 = No peer assessment + average achieving; 3 = No peer assessment + high achieving; 4 = Peer assessment + low achieving; 5 = Peer assessment + average achieving; 6 = Peer assessment + high achieving; Dependent variable: students' performance on a lesson plan project.

Discussion

Aimed at examining if and how students at different learning levels would benefit from peer assessment, this study found that students who conducted peer assessment performed significantly better than students who did not conduct peer assessment on their lesson plan project. Data analysis also revealed a joint influence of peer assessment and learning levels. The effect of peer assessment on students' project performance appeared to depend on student learning levels.

The findings of this study are in line with previous research (Li 2012), in that students of different learning abilities may be affected differently by peer assessment. Despite the students in the peer assessment and no peer assessment groups starting off equally in terms of their performance on the draft lesson plan, LA and AA students in the peer assessment performed significantly better on the lesson plan project than their counterparts in the no peer assessment group. Although the performance of high achievers in the peer assessment group, in numerical values, was higher than that of their counterparts in the no peer assessment group, data analysis did not reveal any significant difference. Further, an intriguing side-finding revealed by Games–Howell *post hoc* tests is that the performance of LA students in the peer-assessment group improved so much that the mean score of their lesson plans, and that of AA students in the no peer assessment, showed no significant difference ($p = .970$). This suggests that peer assessment improved the performance of the LA students in the peer assessment group to attain a level comparable to that of the AA students in the no peer assessment group. No similar finding was discovered for the cell-means comparison for the AA students in the peer assessment group and HA students in the no peer assessment group. All the evidence seems to imply that the immediate learning benefit of peer assessment, as measured by students' project scores, was most impressive for LA students, followed by AA and HA students, respectively.

The positive impact of peer assessment on performance of low and average achievers may be explained, in part, by the social constructivism concepts of the More Knowledgeable Other and the Zone of Proximal Development (Vygotsky 1978). HA students are the more knowledgeable and capable 'others'; their interaction with LA or AA peers in a peer assessment activity may help the latter advance their knowledge, thus reaching their 'proximal development zone'. This interpretation seems to be supported by previous studies on group work. Studies conducted by Webb and colleagues (e.g. Webb 1997; Webb et al. 1998) revealed that group collaboration had differing impacts on students with various learning levels. They further concluded that below-average students seemed to gain great learning advantages when they had opportunities to work with above-average students. Group work tends to benefit below-average students (Roberts 2004). Nevertheless, although this interesting finding of peer assessment's more positive influence on LA students helps disentangle the complexity of peer assessment effect on student learning, the functional significance and sensitivity of the influence remains to be further investigated.

Perusal of data indicated that HA students had the least immediate learning gain in terms of improved scores on their lesson plan project. There may be two possibilities that can be used to explain the finding. First, high-achievers' scores were close to the maximum score on the lesson plan project, which may have created a ceiling effect and decreased the likelihood that the rubric had accurately measured the learning gains of high achievers. In other words, high achievers may simply have not had enough room to demonstrate their improvement. Second, as discussed earlier in this paper, a demonstration of learning may take a wide variety of forms. As a result, a comprehensive assessment of learning must be approached from multiple directions. This study measured students' project performance right after they participated in peer assessment; the findings of the study should not be used to interpret the impact of peer assessment on other types of learning. Topping (1998) argues that the benefits of peer assessment for students may include different domains such as 'cognition and metacognition, affect, social and transferrable skills, and systemic benefits' (254); and in some domains, the advantages of peer assessment 'may accrue before, during or after' (256) its integration. While peer assessment in this study showed great positive influence on project performance of LA and AA students, would the strategy have deferred effects on high-achieving students? Or, while high achievers do not show as much project improvement as low and average achievers after peer assessment, could they have learning gains in other domains such as promoted self-assessment skills, teamwork skills, and communication and interpersonal skills? The researchers suggest that future studies examine the possible impact of peer assessment on students, especially HA students, in other domains and in other types of learning.

The online peer assessment model employed in this study promoted students' performance on the lesson plan project, which coincides with the previous research that well designed and implemented peer assessment may foster student learning (Li 2012; Li, Liu, and Steckelberg 2010). Among the critical features of effective peer assessment captured in the model, the researchers would like to highlight two attributes that have not been sufficiently covered and discussed by previous studies. The first one is clearly defined aims and expectations. This study had the instructor explain the purpose of the activity and explicate the expectations for the students before they were engaged in peer assessment. Students are often reluctant to participate in peer assessment and show little willingness to rate peers. They may question if the learning benefits from peer assessment would outweigh the time and effort they devote to the process (Hanrahan and Isaacs 2001; Topping et al. 2000). The unwillingness and uncertainty may create some issues: assessors are not willing to spend time to review peers' work, and assessees do not take feedback seriously (Hanrahan and Isaacs 2001).

Explicating aims and expectations can considerably encourage students' participation and commitment. Another crucial trait whose significance was largely underestimated was the quality of peer feedback received. Even with effective assessment training, the quality of peer review may vary (Li 2012), largely due to uneven qualifications of reviewers. It is important to educate students that, although they need to consider all feedback, they should evaluate the quality of comments received and only follow those they agree with and deem accurate. The practice of gauging the value of peer feedback is also a learning process, which strengthens students' understanding of marking criteria and quality performance. The researchers believe that these two features contributed greatly to the success of the current project and can be considered by other formative peer assessment models, especially those models with students new to peer assessment.

To maximise students' learning gains and overcome the limitations of paper-based peer assessment, the self- and peer assessment tool in Blackboard, SPABB, was used to facilitate the whole peer assessment process. The use of SPABB effectively avoided the pitfalls of paper-based peer assessment. Anonymity was easily achieved, which greatly reduced peer pressure. Lesson plan submission and peer assessment were conducted entirely online; students had the opportunity to complete these activities at their own pace within the specified time range. Data were managed by the database-driven course management system. The administrative workload was minimal; and the instructor had direct control over the whole process. The advantages of utilising SPABB over paper-based peer assessment were apparent.

Conclusion

This study examined the possible benefits of peer assessment on students of different learning levels. The evaluation results elicit three relevant findings: (1) the online peer assessment model used in this study was effective in promoting students' performance in a lesson plan project; (2) the online peer assessment model had different impact on low achievers, average achievers and high achievers; and (3) low achievers had the greatest benefits from participating in the online peer assessment, followed by average and high achievers.

Significance

The significance of the study lies in it being one of the first few that used an empirical design to investigate the interplay between the impact of peer assessment and student achievement levels. Although general benefits of peer assessment on student learning have been widely documented, research scrutinising if the effect of peer assessment may vary for low, average or high achievers is scarce. The existing literature is limited to mainly descriptive or observational analysis. The findings of the study have important implications for researchers and practitioners who may wish to use peer assessment in their research or daily education practices. Researchers are encouraged to further explore the relationship between peer assessment impact and student learning levels, especially the impact of peer assessment on high achievers. Furthermore, relevant considerations should be taken into account when designing future studies that integrate peer assessment. Practitioners need to know what an effective peer assessment model looks like and how they can use it successfully in their classrooms.

Limitations

Despite the significance of the study, the researchers would like to address two limitations. First, this study used a quasi-experiment design; teacher education students of a technology application course joined either the experimental group (peer assessment) or control group (no peer assessment) based on the class sections they enrolled in. Although a comparison of students' initial performance on the lesson plan project showed no significant difference for the two groups that, to some extent, suggested these two groups were similar at the beginning of the study, the estimation of impact of peer assessment was still subject to possible contamination by confounding variables. Future studies with random selections are warranted.

Second, as discussed above, this study employed a specific group of participants (teacher education students) at a Midwestern US university who might be more open and receptive to different types of learning. Further, in this study, a specific peer assessment model was utilised as the treatment variable. Therefore, the findings of the study may not be generalised to other peer assessment studies or other populations. Future studies should determine if similar findings could be replicated in other populations or with different peer assessment models. The third limitation of the study was the relatively small sample size for the between-subjects 2×3 factorial design. Caution should be taken when generalising these findings to the broader community. Future studies with larger sample sizes are needed to overcome this limitation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Lan Li is an associate professor of Classroom Technology at Bowling Green State University, USA and hold a PhD in Educational Studies from University of Nebraska_Lincoln, USA. Her research interests include technology-facilitated teaching and learning, STEM education and online education. She is enthusiastic about integrating her teaching and research interests with service opportunities that serve the educational needs of surrounding communities. She has published in journals such as *British Journal of Educational Technology*, and *Assessment and evaluation in Higher Education.*

Fei Gao is an assistant professor in Learning Design at Bowling Green State University. She holds a PhD in Educational Psychology and Educational Technology from Michigan State University. Her research focuses on designing technology-enhanced learning environments, computer-mediated communication, e-learning, social media and informal learning. She has published in journals such as *British Journal of Educational Technology* and *The Internet & Higher Education*.

References

ANOVA Options. n.d. http://pic.dhe.ibm.com/infocenter/rentrpt/v1r0m1/index.jsp?topic=%2Fcom.ibm.swg.ba.cognos.ug_cr_rptstd.10.1.1.doc%2Fc_id_obj_anova.html.

Bandura, A. 1977. *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Bauman, E. B. 2012. *Game-Based Teaching and Simulation in Nursing & Healthcare*. New York: Springer.

Brown, D. 2004. *Language Assessment: Principles and Classroom Practice*. New York: Longman.

Brown, S., C. Rust, and G. Gibbs. 1994. *Strategies for Diversifying Assessment in Higher Education*. Oxford: Oxford Centre for Staff and Learning Development.

Brutus, S., M. B. Donia, and S. Ronen. 2013. “Can Business Students Learn to Evaluate Better? Evidence from Repeated Exposure to a Peer-evaluation System.” *Academy of Management Learning & Education* 12 (1): 18–31.
Chen, C. H. 2010. “The Implementation and Evaluation of a Mobile Self-and Peer-assessment System.” *Computers & Education* 55 (1): 229–236.
Cheng, W., and M. Warren. 2005. “Peer Assessment of Language Proficiency.” *Language Testing* 22 (1): 93–121.
Davies, P. 2000. “Computerized Peer Assessment.” *Innovations in Education and Training International* 37 (4): 346–355.
Dochy, F., M. Segers, and D. M. A. Sluijsmans. 1999. “The Use of Self-, Peer-, and Co-assessment in Higher Education: A Review.” *Studies in Higher Education* 24 (3): 331–350.
Espinoza, L., and A. Winsler. 2005. “Zone of Proximal Development (ZPD).” In *Encyclopedia of Human Development*, edited by N. Salkind, 1344–1346. Thousand Oaks, CA: Sage.
Falchikov, N. 2007. “The Place of Peers in Learning and Assessment.” In *Rethinking Assessment in Higher Education: Learning for the Longer Term*, edited by D. Boud and N. Falchikov, 128–143. New York: Routledge.
Falchikov, N., and D. Boud. 1989. “Student Self-assessment in Higher Education: A Meta-Analysis.” *Review of Educational Research* 59 (4): 395–430.
Freeman, M. 1995. “Peer Assessment by Groups of Group Work.” *Assessment & Evaluation in Higher Education* 20 (3): 289–300.
Gibbs, G. 1999. “Using Assessment Strategically to Change the Way Students Learn.” In *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, edited by S. Brown and A. Glasner, 41–53. Buckingham: Open University Press.
Hanrahan, S., and G. Isaacs. 2001. “Assessing Self- and Peer-assessment: The Students’ View’s.” *Higher Education Research and Development* 20 (1): 53–70.
Hersam, M. C., M. Luna, and G. Light. 2004. “Implementation of Interdisciplinary Group Learning and Peer Assessment in a Nanotechnology Engineering Course.” *Journal of Engineering Education* 93 (1): 49–57.
Hiltz, S. R., and B. Wellman. 1997. “Asynchronous Learning Networks as a Virtual Classroom.” *Communications of the ACM* 40 (9): 44–49.
Keppel, G., and T. D. Wickens. 2004. *Design and Analysis: A Researcher’s Handbook*. 4th ed. Upper Saddle River, NJ: Pearson.
Kwan, K., and R. Leung. 1996. “Tutor versus Peer Group Assessment of Student Performance in a Simulation Training Exercise.” *Assessment & Evaluation in Higher Education* 21 (3): 205–214.
Li, L. (2012). How do Students of Diverse Achievement Levels Benefit from Peer Assessment? *International Journal for the Scholarship of Teaching & Learning* 5 (2). http://digitalcommons.georgiasouthern.edu/ij-sotl/vol5/iss2/14.
Li, L., X. Liu, and A. L. Steckelberg. 2010. “Assessor or Assessee: How Student Learning Improves by Giving and Receiving Peer Feedback.” *British Journal of Educational Technology* 41 (3): 525–536.
Li, L., X. Liu, and Y. Zhou. 2012. “Give and Take: A Re-analysis of Assessor and Assessee’s Roles in Technology-facilitated Peer Assessment.” *British Journal of Educational Technology* 43 (3): 376–384.
Li, L., and A. L. Steckelberg. 2004. “Using Peer Feedback to Enhance Student Meaningful Learning.” In *Proceedings of the Association for Educational Communications and Technology International Conference*, 576–587. Chicago, IL.
Li, L., and A. L. Steckelberg. 2005. “Impact of Technology-mediated Peer Assessment on Student Project Quality.” In *Proceedings of the Association for Educational Communications and Technology International Conference*, 307–313. Orlando, FL.
Li, L., A. L. Steckelberg, and S., Srinivasan. 2008. “Utilizing Peer Interactions to Promote Learning through a Web-based Peer Assessment System.” *Canadian Journal of Learning and Technology* 34 (2). http://www.cjlt.ca/index.php/cjlt/article/view/497/228.
Liu, X., and L. Li. 2013. “Assessment Training Effects on Student Assessment Skills and Task Performance in a Technology-facilitated Peer Assessment.” *Assessment & Evaluation in Higher Education* 39 (3): 275–292.

Maclellan, E., and R. Soden. 2008. "Successful Learners, Confident Individuals, Responsible Citizens and Effective Contributors to Society: Exploring the Nature of Learning and its Implications in Curriculum for Excellence." *Scottish Educational Review* 40 (2): 29–37.

McDonald, R., D. Boud, J. Francis, and A. Gonczi. 1995. *New Perspectives on Assessment*. Paris: UNESCO.

Mehrens, W. A., W. J. Popham, and J. M. Ryan. 1998. "How to Prepare Students for Performance Assessments." *Educational Measurement: Issues and Practice* 17 (1): 18–22.

Orsmond, P., S. Merry, and K. Reiling. 1996. "The Importance of Marking Criteria in the Use of Peer Assessment." *Assessment & Evaluation in Higher Education* 21 (3): 239–250.

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.

Race, P. 2007. *The Lecturer's Toolkit*. 3rd ed. London: Routledge.

Roberts, T. S. 2004. *Online Collaborative Learning*. Hershey, PA: Information Science Publishing.

Robinson, J. 1999. "Computer-assisted Peer Review." In *Computer-assisted Assessment in Higher Education*, edited by S. Brown, J. Bull, and P. Race, 95–102. London: Kogan Page.

Somervell, H. 1993. "Issues in Assessment, Enterprise and Higher Education: The Case for Self-, Peer and Collaborative Assessment." *Assessment & Evaluation in Higher Education* 18 (3): 221–233.

Sung, Y. T., K. E. Chang, S. K. Chiou, and H. T. Hou. 2005. "The Design and Application of a Web-based Self-and Peer-assessment System." *Computers & Education* 45 (2): 187–202.

Topping, K. 1998. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 68 (3): 249–276.

Topping, K., E. F. Smith, I. Swanson, and A. Elliot. 2000. "Formative Peer Assessment of Academic Writing between Postgraduate Students." *Assessment & Evaluation in Higher Education* 25 (2): 149–169.

Van Der Stuyf, R. 2002. *Scaffolding as a Teaching Strategy.* Adolescent Learning and Development. Section 0500A. condor.admin.ccny.cuny.edu/~group4/Van%20Der%20Stuyf/Van%20Der%20Stuyf%20Paper.doc.

Van Zundert, M., D. Sluijsmans, and J. Van Merriënboer. 2010. "Effective Peer Assessment Processes: Research Findings and Future Directions." *Learning and Instruction* 20 (4): 270–279.

Violato, C., and J. Lockyer. 2006. "Self and Peer Assessment of Pediatricians, Psychiatrists and Medicine Specialists: Implications for Self-directed Learning." *Advances in Health Sciences Education* 11 (3): 235–244.

Vygotsky, L. S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wang, Y., H. Li, Y. Feng, Y. Jiang, and Y. Liu. 2012. "Assessment of Programming Language Learning based on Peer Code Review Model: Implementation and Experience Report." *Computers & Education* 59 (2): 412–422.

Webb, N. M. 1997. "Assessing Students in Small Collaborative Groups." *Theory Into Practice* 36 (4): 205–213.

Webb, N. M., K. M. Nemer, A. W. Chizhik, and B. Sugrue. 1998. "Equity Issues in Collaborative Group Assessment: Group Composition and Performance." *American Educational Research Journal* 35 (4): 607–651.

Zhao, Y. 1998. "The Effects of Anonymity on Computer-mediated Peer Review." *International Journal of Educational Telecommunications* 4 (4): 311–345.