The Effects of Social Messaging on Students' Learning and Intrinsic Motivation in Peer Assessment

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

Fei Gao
Bowling Green State University, gaof@bgsu.edu

Shuqing Guo
Xi'an International Studies University

The Effects of Social Messaging on Students' Learning and Intrinsic Motivation in Peer Assessment

Lan Li
Fei Gao
Bowling Green State University

Shuqing Guo
Xi'an International Studies University

Abstract

This study aims to gain a better understanding of how the newly arisen social messaging may impact the practice of peer assessment. Seventy-nine ESL (English as Second Language) students reviewed each other's English essays in three peer assessment groups: a three-member group using wiki (wiki group), a three-member group using social messaging (small messaging group), and a six-member group using social messaging (big messaging group). Data analysis suggested that peer assessment facilitated by social messaging can be at least of the same effectiveness as wiki-facilitated peer assessment on ESL students' writing skills and intrinsic motivation. In addition, the findings indicated that students in the small messaging group outperformed students in the big messaging group on essay writing, and reported a significantly higher rating on Perceived Competence, a positive indicator of the behavioral measures of intrinsic motivation, than students in the big messaging group.

Key Words: peer assessment, social messaging, student learning, intrinsic motivation, messaging group size

Overview of Formative Peer Assessment

Peer assessment is defined as a process in which students judge peers' work or performance based on agreed benchmarks (Falchikov, 2007). Despite many variants, peer assessment distinguishes three main types: formative, summative, or a combination of both. The intention of formative peer assessment is to engage students in both roles as assessor and assessee and to facilitate peer learning. During the process, as assessors, students evaluate the quality of peers' work and provide constructive feedback. As assessees, students gauge the value of peer feedback they receive and improve their own work accordingly (Li, Liu & Steckelberg, 2010; Li, Liu & Zhou, 2012).

Peer assessment has been widely applied across various disciplines, such as medicine, education, computer science, engineering, etc. (Li & Gao, 2016), and used in various educational settings from elementary schools through post-graduate programs including with students with special needs (Scruggs & Mastropieri, 1998). A Google Scholar search of "Peer Assessment" publications since the year of 2000 generated almost 2 million entries. A growing body of research suggests

that well-designed and properly implemented peer assessment may positively impact a range of outcome variables including students' academic achievements and cognitive development. Commonly reported benefits include improved learning outcome (Liu & Li, 2014, Li & Steckelberg, 2005; Pope, 2001); enhanced critical thinking skills and self-regulation (Harrison, O'Hara, & McNamara, 2015); promoted motivation, autonomy, and responsibility (Pope, 2001; Somervell, 1993); increased access to timeliness of feedback (Gibbs, 1999) and quantity of feedback (Topping, 2009); boosted student interaction (Peng, 2010) and interpersonal skills (Brown, Topping, Henington, & Skinner, 1999); and enhanced understanding of assessment criteria (Burke & Pieterick, 2010). Among the many potentials reported, one of the most prominent is that peer assessment encourages active and reflective learning, and empowers students to take control of their own learning growth (Li et al., 2010; Li et al., 2012). These are essential life skills that students need to succeed in their education, career, and life (Li, 2017, Topping, 2009).

The Role of Technology in Peer Assessment

Although peer assessment has been around for more than three centuries (Shema, 2014), the practice was mainly paper-based and did not integrate much digital technology until the end of the last century. One of the earlier technology-facilitated models was reported by Rushton and his team (Rada, Acquah, Baker & Ramsey, 1993; Rushton, Ramsey & Rada, 1993) in the 1990s. The multi-user hypermedia system called MUCH was used at the University of Liverpool, England, to support students' collaborative learning. One feature of MUCH was specifically designed to facilitate peer assessment. The system allowed students to read and critique each other's work. One drawback, as noted by the authors, was that anonymity was not provided during the process.

At the beginning of the 21st century, the ease of access to computer hardware and software, and improved online capabilities drove an increased interest in the joint value of technology and assessment. The literature witnessed a proliferation of studies that investigated the impact of various technology-supported peer assessment models on different aspects of students' learning process. For example, a few studies (e.g., Liu, Lin, Chiu & Yuan, 2001; Liu , Lin, & Yuan, 2002) portrayed a network that supported peer assessment and promoted students' critical thinking skills. After students uploaded their work to the network, it allowed students to perform both roles as assessor and assessee. Students were able to review and provide feedback to peers' work as assessors. Furthermore, they were able to view feedback provided by peers and improve their own work as assessors. Data analysis suggested that this process had a positive impact on the quality of students' work. Davies (2000) reported a system called CAP (Computerized Assessment with Plagiarism) that was used in a second-year undergraduate module at the University of Glamorgan to support peer review and minimize the risk of plagiarism. Findings indicated that engaging students in reviewing each other's work through CAP facilitated learning and effectively addressed the issue of plagiarism. While a sizeable number of studies employ these specifically designed systems to facilitate peer interaction, many educators and researchers turn to added alternative assessment features in Learning Management Systems, (LMS) such as Blackboard or Moodle, to support their peer assessment activities (Li, 2017, Li &

Gao, 2016, Lu, Warren, Jermaine, Chaudhuri & Rixner, 2015, Shen & Huang, 2006).

In no small part, the advancement of technology has shaped and reshaped the practice of peer assessment. With the maturation and diffusion of web 2.0 as social technology, platforms that facilitate social interaction, such as blog, wiki, and podcasting, began to emerge at the turn of the century. When YouTube, Facebook, and Twitter became available between 2004 and 2006, social media began to explode in popularity. Such networks soon became the most visited sites for young people (Shih, 2011). The potential of these podiums as instructional tools has drawn considerable and growing attention from all educational sectors (Alexander, 2006), and peer assessment is no exception. A growing number of peer assessment studies that employed these platforms have been reported since then. Wiki, due to its inborn feature that enables collaborative content editing, soon became one of the most commonly utilized tools reported in these studies (e.g. De Wever, Van Keer, Schellens, & Valcke, 2011, Lin & Yang, 2011, Xiao & Lucking, 2008), a good number of which were conducted in the field of English language and academic writing. For example, Alshumaimeri (2011) conducted a study to examine the effects of a wiki-based peer assessment approach on EFL (English as Foreign Language) students' writing skills. The control group received traditional instruction, while the experimental group used wiki-based peer assessment. The authors found that students in the experimental group outperformed those in the control group in writing accuracy and quality. In addition to the use of wiki, Facebook and Twitter have also been often used to facilitate peer review. Shih (2011) examined the influence of Facebook-mediated peer assessment on students' English writing skills and students' attitudes. Twenty-three college students at a technological university in Taiwan posted their English writing assignments on Facebook, reviewed each other's submissions, and provided comments on the platform weekly for seven writing tasks. The findings of the study suggested that students, regardless of their initial English proficiency levels, had significantly improved their English writing skills at the end of the study. In addition, the Facebook-integrated peer assessment had a positive impact on students' attitudes and perceptions. Luo, Dani & Cheng (2016) conducted a case study that employed Twitter as a backchannel to support a peer-teaching activity in a face-to-face early childhood science method course at a Midwestern U.S. university. Students observed peer teaching and provided feedback by two means: with paper and pencil or through Twitter. A comparison of the two types of feedback suggested that Twitter-supported peer feedback was at least as effective as the paper-based approach. Nevertheless, the analysis of students' qualitative survey data suggested that students exhibited some degree of skepticism toward the value of the activity. Overall, the majority of these studies seem to suggest that social media integration in peer assessment may engage students and favorably impact their learning.

Recently, the rise of social messaging signals a new shift in the landscape of social media. The characteristics of these social messaging apps include, but are not limited to, one-to-one or group chats, video or voice calls, file sharing, status updates, push notification, and payments. Many social activities are no longer happening in public platforms like Facebook and Twitter. Instead they are moving to smaller, more private settings. As Read (2016., para. 1) stated, "As we progress through 2016, and beyond we'll start to notice most social activity is no longer going to

happen in public, instead transitioning to private groups and messaging apps." This move appears massive. After analyzing the use data for the four biggest social messaging apps (WhatsApp, Messenger, WeChat, and Viber) and the four social networking giants (Facebook, Instagram, Twitter, and LinkedIn), Business Inside intelligence (2016) announced that use of the four messaging apps has surpassed that of the four social networks. Read calls this "the biggest Internet phenomenon since the App Store." As the group messaging paradigm starts to populate and rule the social media realm, education faces a new wave of questions, challenges, and also opportunities. How can we harness the power of this new norm of social interaction to reach and engage students? What does this mean in the area of peer assessment? Since the early appearance of technology integration in peer assessment in the 1990s, technological advances have always been transforming the practice in this field. The researchers of this study were interested in examining what this emerging paradigm shift in social media may offer to researchers in the area of online peer assessment.

Since this movement is rather new, literature review reveals that reported research integrating social messaging platforms to support peer assessment is scarce. Among the few scattering studies, one (Wu, Hou, & Hwang, 2012) used MSN instant messaging system to facilitate peer assessment. This study explored the interactive process of synchronous peer assessment activities via MSN but did not focus on the "social" aspect, as students participated in the peer assessment process on a one-to-one basis. With WhatsApp getting more popular in the past five years, it has been used in a couple of peer assessment studies. For example, Güler (2017) explored the use of WhatsApp to support anonymous and non-anonymous peer assessment. Eighty-four college students were assigned into an anonymous group (WhatsApp personal chat) and a non-anonymous group (WhatsApp group chat) and reviewed each other's work. Güler concluded that WhatsApp as a messaging tool can be used as an effective apparatus to support peer assessment. He further reported that anonymity did not have a significant impact on students' attitudes, as both groups in his study had pretty positive attitudes toward peer assessment. Fattah (2015) examined the effectiveness of WhatsApp as a mobile learning podium to cultivate college students' writing skills. Students were assigned into a control group and an experimental group. Students in the control group followed the traditional approach to develop their writing skills. Students in the experimental group developed and shared their writing drafts in a WhatsApp group chat, and then reviewed each other's work and corrected errors. The results of the study suggested that students in the experimental (WhatsApp) group outperformed those in the control group on their writing evaluation. As the above handful studies show, research on social messaging and peer assessment is still in its early stage. The applications and impact of social messaging apps on peer assessment are still to be learned. Given how little is known, the researchers were interested in seeing how the new messaging platform may function differently than older social media platform—wiki—when integrated in peer assessment activities. As the exploration of the use of social messaging in peer assessment is complex and multifaceted, the researchers decided to conduct a series of two studies with different focuses. The current paper reports the first study, which aimed to compare the potential impact of peer assessment operated on wiki and social messaging platforms on students' learning and their intrinsic motivation in variously sized online groups. The second

paper of the series intends to explore the impact of social messaging on dimensions of students' online behavior and interaction.

Optimal Group Size in Peer Assessment

The literature does not present a consensus regarding the optimal group size for effective and meaningful group interaction. Some researchers suggest that the optimal group size be moderate such as five (Fay, Garrod, & Carletta, 2000) or seven (Blenko, Mankins, & Rogers, 2010) for leaderless groups. Nevertheless, others express their preferences to bigger groups. Roger and Link (1996) assert that "the number of possible social interactions begins to explode in groups with more than 5 people" (para. 5). Furthermore, Hashmi (2017) reported that larger group settings, in general, generate more collective intelligence than smaller group settings. The same inconclusive results regarding the optimal group size are reported in the area of peer assessment. While some studies validate smaller groups, others endorse otherwise. For example, Valacich, Dennis, and Nunamaker (1992) stated that students in larger peer review groups (nine members) provided more and better ideas than those in smaller peer review groups (three members) on group idea generating. However, other studies seem to suggest that smaller peer review group such as pairs would likely to create and sustain effective learning environments (e.g. Bennett, Parker, & Smigiel, 2012), and encourage students' participation (e.g. Hung, Chen, & Samuelson, 2016). The interplay between group size, students' participation and performance in peer assessment is seemingly not well understood. The reciprocal relationships between these factors have become more complicated with the advent of the Internet, and all the opportunities and challenges that come along with it.

In previous studies conducted by the researchers (e.g. Li, 2017, Li, 2018, Li et al., 2010, Li et al., 2012), peer assessment groups usually consisted of three members. Each member in the same group would review the performance of the other two group members. While the group size appeared functional, some students complained that the amount and value of peer feedback they received was limited. The issue seemed the most bothersome when both assessors of a student's work had struggles providing constructive comments. Hence, the researchers wondered if a big group setting (with six members) supported by the social features of a messaging app would remedy the problem.

Intrinsic Motivation

As compared to extrinsic motivation where the impetus to act originates from external sources such as prizes or awards, Intrinsic motivation refers to an individual's desire to perform an activity or task solely derived from his/her internal drive such as joy and interest (Ryan & Deci, 2000). Studies suggest that intrinsic motivation is one of the key indicators of students' academic performance (Pascoe et al., 2018) and directly predicts motivated behaviors. Buzdar and his colleagues (2017) assert that intrinsically motivated students are more willing to take academic risks and confront challenging tasks. These students also are more likely to exert effort into achieving goals that they have set for themselves.

The Intrinsic Motivation Inventory (IMI), originally developed by Ryan, Mims and Koestner (1983), intends to measure an individual's subjective experience related to a target activity. The current study explicitly focuses on two IMI subscales: Perceived Competence and Pressure. The Perceived Competence subscale is the positive predictor of "both self-report and behavioral measures of intrinsic motivation" (Self-Determination Theory, n.d., para. 1). and the Pressure subscale is theorized to be a negative predictor of intrinsic motivation.

In light of the literature reviewed and the rationales discussed, the study asked the following research questions:

1). Is there a significant difference in students' mean essay scores between students who participate in peer assessment groups that integrate different types of social media (wiki, small messaging group—three students, and big messaging group—six students)?
1a). If so, which groups differ?

2). Is there a significant difference in students' mean Perceived Competence scores between students who participate in peer assessment groups that integrate different types of social media (wiki, small messaging group—three students, and big messaging group—six students)?
2a). If so, which groups differ?

3). Is there a significant difference in students' mean Pressure scores for students who participate in peer assessment groups that integrate different types of social media (wiki, small messaging group—three students, and big messaging group—six students)?
3a). If so, which groups differ?

The researchers hypothesized that there would be a significant difference on essay scores, Perceived Competence Scores, and Pressure scores between students who participated in peer assessment enabled by wiki, small messaging group (SMG), and big messaging group (BMG). Specifically, it was hypothesized that the BMG students would outperform SMG students and those in the wiki group. In addition, the BMG students would report higher mean essay scores and Perceived Competence scores, and lower Pressure scores than SMG students and students in the wiki group, as the big group setting would encourage students' interaction and effectively foster active learning, therefore promoting students' intrinsic motivation.

Methodology

Participants

This study recruited a convenience sample of 79 freshman ESL (English as Second Language) students enrolled in a compulsory English reading and writing course for English majors at a major Chinese language institute. Among reported, the participants had an average age of 19. They consisted of 71 females and 8 males. The majority of participants (92%) were Han nationality. The rest were minorities.

Procedure

As Figure 1 shows, students were assigned to a wiki peer assessment group (control) and two groups with social messaging platform integration (SMG, BMG) based on the class sessions they were enrolled in. Although this study used a typical quasi-experimental design based on pre-defined class sessions, the assignment of students into these class sessions was random, which made it less likely that there were other differences between conditions and therefore minimizing the problem of confounding variables.
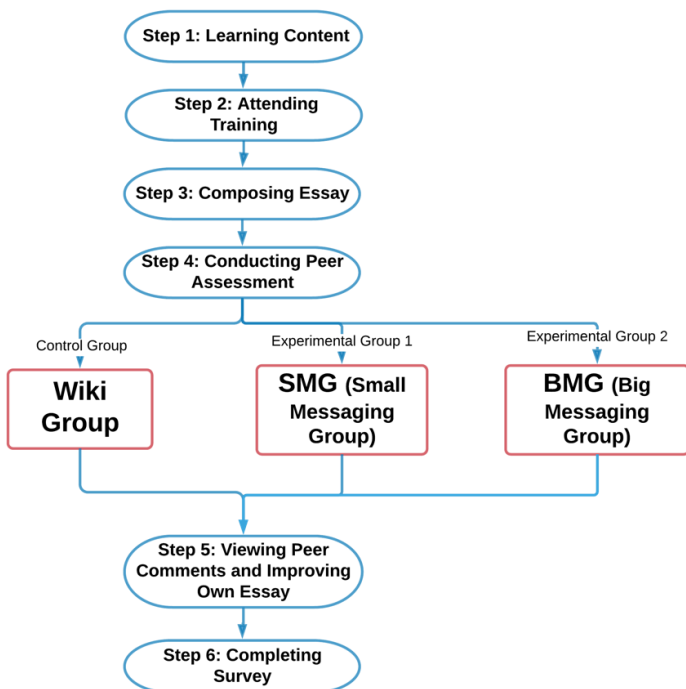

Figure 1. Study Process Flowchart

Step 1: Learning Content. Students in all groups learned the content (argumentation essay). According to the textbook entitled *Ten Steps to Improving College Reading Skills* (Langan, 2008) that the English course used, "a good argument is one in which you make a point and then provide persuasive and logical evidence to back it up" (p.380). The purpose of this assignment was to check students' English proficiency and their ability to think critically.

Step 2: Attending Training. After students learned what the argumentative essay was, they were presented with the grading rubric (see Measures below for detail) so they understood what a quality essay looked like. Two example essays were provided for students to practice their assessment skills. Class discussion was held to explain the rubric, to answer students' questions and to resolve any disagreements between students' ratings and the instructor's ratings of the example projects. Students were also informed that they would participate in a peer assessment

activity, in which they were to review each other's essays based on the rubric and help each other improve their projects.

Step 3: Composing Essay. In this step, students in all groups composed their augmentation essays. Each student first identified a topic, usually a controversial one, to center their essay on. Each student then stated the main point of his/her essay, and supported the main point by providing three supporting details and/or examples.

Step 4: Conducting Peer Assessment. Based on class sessions enrolled, students were assigned into three groups: wiki group (Control), SMG with three members (Experiment 1), and BMG with six members (Experiment 2). The control group followed a typical online peer assessment process. Students uploaded their completed essays to an online system called Shimo, which is a cloud-based collaboration app with wiki features that facilitate online file sharing and editing. Students were assigned into wiki groups with three members. Students in the same group reviewed each other's essays and provided feedback based on the rubric. Students accessed the platform to conduct peer assessment (as assessor) and to view comments provided to their own essays (as assessee). Even though students were encouraged to interact with each other by leaving comments on the review pages, interaction in this platform was limited and mainly one-way.

Students in the two experimental groups used WeChat as the platform to conduct the peer assessment activity. WeChat is a multi-purpose social messaging program that supports instant messaging, group chatting, data transferring, status update, payment, etc. Among its various features, the current study only employed the following functions: group chat, push notification, and file sharing. Multiple chat groups were formed within each experimental group. Settings for both experimental groups were identical except that each chat group in Experimental Group 1 contained three students (SMG), while each chat group in Experimental Group 2 contained six students (BMG). SMG students were instructed to post their essays in their own chat groups and review the other two essays shared in the same chat group. The BMG students had the choice to choose two of the six essays posted in their chat group to review, on the condition that all essays should have at least two reviews. That means that if any essay had received two reviews, group members needed to move on and choose other essays to conduct their reviews.

Unlike the wiki group, interaction in the SMG and the BMG was intense, interminable, and multidirectional. Due to the social media feature of WeChat, students in the same chat group received notifications once new messages were posted. Students had access to all messages posted and files shared in the same chat group. Students appeared more devoted and dedicated in the process. A number of students posted multiple messages around the clock to share their opinions, to argue, or to debate. It is fair to say that more dialogues between students were identified in the SMG and the BMG.

Step 5: Viewing Peer Comments and Improving Own Essay. Students viewed peer feedback provided and improved their own essays in all conditions. It was made explicit that feedback may

come in varied quantity and quality. Students should judge the value of peer feedback before making adoptions. All students submitted their revised essays.

Step 6. Completing Survey. All students completed the post peer assessment survey. The online survey included demographic questions such as age, gender, and nationality. Further, the survey contained two subscales: Perceived Competence and Pressure from the Intrinsic Motivation Inventory (IMI). See Measures below for more details.

Measures

Essay Rubric. The quality of an essay is measured using the essay grading rubric, which consists of three criteria: Main Idea, Supporting Details, and Quality of Writing. Each criterion contains questions that require assessors to make judgment and then provide justifications. For example, under the "Main Idea" category, there are three questions: *Is the main idea clearly presented? Does the main idea convey a generally arguable idea? Is the main idea followed by three supporting details?* For the each of the three questions, assessors were asked to provide a Yes or No answer. Furthermore, they needed to justify their choices and provide suggestions to help assessees improve their writing. While the maximum possible score students could achieve in the project was 10 points, decimal scores were allowed for grading. For example, a student's essay score could be 9.6 out of 10. Following the rubric, both the instructor and an independent grader graded all students' essays with an inter-rater reliability (Pearson correlation) established at $r = .848$. All disagreements were resolved after discussion.

Perceived Competence & Pressure/Tension Subscales. The Intrinsic Motivation Inventory (IMI) consists of 7 subscales: Interest/Enjoyment, Perceived Competence, Effort/Importance, Pressure/Tension, Perceived Choice, Value/Usefulness, and Relatedness. Among the 7 subscales, the researchers in this study were particularly interested in two—Perceived Competence and Pressure/Tension. According to IMI, the Perceived Competence and Pressure/Tension concepts are theorized to be a positive predictor of "self-report and behavioral measures of intrinsic motivation" (Self-Determination Theory, n.d., para. 1). and negative predictor of intrinsic, respectively. Both the subscales are rated on a 7-point Likert scale ranging from 1 (not at all true) to 7 (very true). A higher score on Perceived Competence indicates that the individual feels more competent. A higher score on Pressure/Tension indicates that the individual feels more pressured and tensed. The Perceived Competence subscale consists of seven items measuring students' self-judgment of their own ability to perform the peer assessment task. Example items include "I think I am pretty good at this activity," and "I am satisfied with my performance at this task." The Pressure/Tension subscale consists of five items gauging the level of pressure that participants experience during the peer assessment process. Items include "I felt very tense while doing this activity," and "I was very relaxed in doing these."

IMI has been employed in a number of studies across various disciplines. The reliability and validity tests showed that the subscales were reliable and had factorial validity (McAuley, Duncan & Tammen, 1989). Tests for internal consistency for the two subscales were also quite adequate

in this study, as determined by a Cronbach's alpha of .796 for Perceived Competence and .888 for Pressure/Tension.

Results

A one-way multivariate analysis of variance (MANOVA) was run to determine the effect of different peer assessment venues (wiki, SMG, BMG) on students' essay scores, Perceived Competence scores, and Pressure/Tension scores, as the researchers believed that the treatment would likely affect participants in more than one way.

Preliminary assumption checking suggested that there were four univariate outliers in the data, as suggested by a boxplot. A one-way MANOVA test with and without the outliers was run to determine whether the outliers had an appreciable impact on the analysis. Results suggested that the same conclusions should be drawn from the analysis with the two data sets (with and without outliers). Therefore, the researchers decided to keep the outliers in the data. The Pressure/Tension scores were normally distributed for each group, as assessed by Shapiro-Wilk's test ($p > .05$). The Perceived Competence and essay scores were normally distributed for the wiki group and the SMG as assessed by Shapiro-Wilk's test ($p > .05$), but not the BMG (Shapiro-Wilk's test, $p < .05$). The researchers decided to run the one-way MANOVA regardless, as the test is fairly robust to deviations from normality (Laerd Statistics, 2015, Mertler, & Vannatta, 2005). No multicollinearity was detected, as Pearson correlation was $r = .336$ ($p = .002$) between essay and Perceived Competence scores, $r = .041$ ($p = .722$) between essay scores and Pressure/Tension scores, and $r = -.440$ ($p < .0001$) between Pressure/Tension and Perceived Competence scores. The relationship between Perceived Competence, Pressure/Tension, and essay scores in each group appeared linear for all groups. No multivariate outliers were detected in the data, as assessed by Mahalanobis distance ($p > .001$). There was homogeneity of variance-covariances matrices and homogeneity of variances, as assessed by Box's test of equality of covariance matrices ($p = .938$) and Levene's Test of Homogeneity of Variance ($p > .05$), respectively.

Table 1 demonstrates the mean scores and standard deviations of students' essay scores, Perceived Competence scores, and Pressure/Tension scores. Table 1 shows that both student essay scores and Perceived Competence scores in the SMG, in numerical values, is higher than those of the other two groups.

Table 1
Mean Scores and Standard Deviations of Student essay performance Scores, Perceived Competence Scores, and Pressure/Tension Scores by Groups

	Wiki N=22		Small Messaging Group (SMG) N=29		Big Messaging Group (BMG) N=28	
	M	SD	M	SD	M	SD
Essay	7.05	1.18	7.31	1.27	6.27	1.01

Competence	4.55	0.80	4.83	0.82	4.09	0.74
Pressure	3.01	1.28	2.73	1.35	3.37	1.43

Note: M=Mean, SD=Standard Deviation, N=number

The differences between groups on the three combined dependent variables were statistically significant, F(6, 148) = 3.28, p = .005; Wilks' Λ = .779; partial η2 = .117. Follow-up univariate ANOVAs (Table 2) showed that both essay scores, F(2, 76) = 6.105, p = .003; partial η2 = .138, and Perceived Competence scores, F(2, 76) = 6.360, p = .003; partial η2 = .143, were statistically significantly different between the students from different groups. There was no statistically significant difference in Pressure/Tension scores between the students from different groups.

Table 2
Univariate ANOVA Tests of Dependent Variables (Mean Essay Score, Mean Perceived Competence Score, and Mean Pressure/Tension Score)

Dependent Variable	df	df error	F	Sig	Mean Square	Partial Eta Squared	Observed Power
Essay	2	76	6.105	.003	8.226	.138	.876
Perceived Competence	2	76	6.360	.003	3.934	.143	.889
Pressure/Tension	2	76	1.586	.212	2.932	.040	.326

Data are expressed as mean ± standard deviation. As Table 1 shows, the mean essay score increased from the BMG (6.27 ± 1.01), to the wiki group (7.05 ± 1.18), to SMG (7.31 ± 1.27), in that order. Tukey post hoc analysis revealed that the increase from the BMG to the SMG, 1.04, 95% CI (0.31 to 1.78), was statistically significant (p = .003), but no other group differences were found statistically significant for mean essay score.

Mean Perceived Competence scores increased from the BMG (4.09 ± 0.74), to the wiki group (4.55 ± 0.80), to SMG (4.83 ± 0.82), in that order. Tukey post hoc analysis revealed that the increase from the BMG to the SMG, 0.74, 95% CI (0.24 to 1.24) was statistically significant (p = .002), but no other group differences were found statistically significant for the mean Perceived Competence score.

Discussion and Conclusion

Results showed that for all three measures (essay score, Perceived Competence score, and Pressure/Tension score), the SMG had either the largest or smallest value, then followed by the wiki group, and then followed by the BMG. Furthermore, the SMG students significantly outperformed the BMG students on the essay quality and had significant higher Perceived Competence mean score than the BMG students.

One hypothesis of the study was that the BMG students would manifest better writing skills than the SMG students, as it was anticipated that the former setting would engage students more and therefore generate more collective intelligence and facilitate a more interactive and dynamic learning environment. On the contrary, the findings of this study revealed that, instead, the SMG students outperformed students in the big group setting on essay writing. This finding was quite intriguing and surprising. Based on the findings of previous research (see Optimal Group Size in Peer Assessment), it seemed reasonable to predict the superiority of a BMG with six group members over a SMG with three group members on students' writing skills. Interestingly, however, our findings appeared to suggest otherwise. What's more, it appeared that the SMG students also demonstrated better intrinsic motivation than the BMG students. As the data analysis showed, the SMG students had a significantly higher mean score of Perceived Competence (positive indicator of the behavioral measures of intrinsic motivation) than those in the BMG group. Although the comparison of students' Pressure/Tension mean scores (the negative indicator of intrinsic motivation) between the SMG and the BMG did not show statistically significant difference, the SMG students did report a lower pressure score as compared to that reported by the BMG students (2.73 vs 3.37, respectively), which indicated a higher intrinsic motivation in the SMG setting.

The comparisons of the three measures between the SMG and the BMG seemed to suggest that social messaging may pose better learning and motivational outcomes in small settings than big settings. There were many potential explanations of why these interesting results in this study were observed. 1) Many of the optimal group size studies occurred before 2000 or earlier and in face-to-face settings. What was identified as favorable group sizes may not be true for online environments. Future studies should be conducted to identify optimal group sizes for online collaborations, specifically in online peer assessment. 2) Online collaboration also takes various forms and may use various collaboration tools. As use of collaboration tools in group settings may contribute to group collective intelligence (Hashmi, 2017), future studies should compare uses of various social media podiums in peer assessment to see how they may function differently to engage students and foster learning in peer assessment. 3) Learning is viewed as a complicated and multifaceted process that may be influenced and triggered by various factors such as "the nature of the task, individual learner differences, learners' beliefs and attitudes, and learning contexts" (Li, 2018, p. 9). This study measured one learning outcome (student essay score) and one motivational outcome (intrinsic motivation as represented by its positive and negative indicators, Perceived Competence and Pressure/Tension, respectively) immediately after the peer assessment. The superiority that the SMG setting over the BMG setting demonstrated on these measures does not necessarily decline the latter's value and impact on other aspects of learning. A deeper analysis of the results by task level and by consideration of learner proficiency and learning context should be conducted to examine if and how these factors may have contributed to the observations of current results. Future experiments that look into students' long-term or delayed learning and other dimensions of motivational outcome such as self-regulation and self-efficacy are warranted. In the second study of the series, the researchers aim to analyze and compare the group dynamics in both BMG and SMG settings in

order to identify possibly different attitudinal and behavioral characteristics of these two settings.

One finding the researchers would like to highlight was that it was feasible to employ newly emerged social messaging tools to support online peer assessment. The results in this study indicated there were no significant differences between the wiki group and the BMG, and between the wiki group and the SMG on all three measures. The results of these comparisons suggested that social messaging tools, when properly implemented, may foster online peer assessment as effectively as the commonly used wiki, if not better.

The current study has significant implications for advancing our knowledge and understanding of using social media as a learning tool. In the digital age, students are immersed in social media, and their life revolves around social media. As a result, it is imperative to engage them in social media learning (Shehu & Besimi, 2017). This is also true in the field of peer assessment. The development and deployment of peer assessment has constantly evolved over the last two decades, as swiftly advanced technology is drastically reshaping the world and education. While affordance of these Web 2.0 tools has allowed for creativity and innovations when engaging students in peer instruction activities, it is critical to understand how different generations of technology may engage students and foster their learning in different ways and offer different promises. As this is one of the first few studies that investigate the integration of social messaging with peer assessment, the researchers hope the findings of the study could shed light on studies that examine the influence of the new wave of social media on students learning, especially in online collaborative learning environments.

The present findings should be interpreted with a few limitations taken in consideration. First, participants in the current study were freshman EFL students who were enrolled in a compulsory English reading and writing course for English majors at a Chinese language institute. The sample was drawn because of its easy accessibility and ready availability, therefore, does not represent the whole population. For instance, this group joined the study may possess different characteristics as compared with other groups learning in their native languages. Interpretation of the findings should not be readily generalized to a broader or a different population. Future studies should use samples consisting participants from other backgrounds, regions, or ethnic groups. Second, a quasi-experiment design was used in this study due to ethical and practical considerations. Students were assigned into either the wiki group, the SMG, or the BMG based on class sessions they enrolled in. Although this current study did not use random assignment, students were assigned into their parallel class sessions indiscriminately. This condition may have minimized group differences. However, the researchers still cannot totally rule out possible confounding variables. The researchers suggest that future studies should replicate the design with random assignments. Third, this study only provided one snapshot of students' writing skills and their intrinsic motivation immediately after the peer assessment intervention. In recognition of the complex and multifaceted nature of this study, future studies are warranted to explore other aspects of students' learning.

Data Availability

Due to ethical restrictions, data are available upon request and approval by the Institutional Review Board or equivalent from the Institution where data were collected.

References

Alexander, B. (2006). Web 2.0: A new wave of innovation for teaching and learning? *EDUCAUSE Review*, 41(2), 32–44.

Alshumaimeri, Y. (2011). The effects of wikis on foreign language students writing performance. *Procedia-social and behavioral sciences*, 28, 755-763.

Bennett, P. N., Parker, S., & Smigiel, H. (2012). Paired peer review of university classroom teaching in a school of nursing and midwifery. *Nurse education today*, *32*(6), 665-668

Business Inside Intelligence (2016, September 20). Messaging apps are now bigger than social networks. Retrieved from https://www.businessinsider.com/the-messaging-app-report-2015-11?IR=T

Buzdar, M. A., Mohsin, M. N., Akbar, R., & Mohammad, N. (2017). Students' academic performance and its relationship with their intrinsic and extrinsic motivation. *Journal of Educational Research*, *20*(1), 74-82.

Blenko, M. W., Mankins, M. C., & Rogers, P. (2010). *Decide & deliver: 5 steps to breakthrough performance in your organization*. Harvard Business Press.

Brown, C. C., Topping, K. J., Henington, C., & Skinner, C. H. (1999). Peer monitoring of learning behaviour: The case of 'Checking Chums'. *Educational Psychology in Practice*, 15(3), 174-182.

Burke, D., & Pieterick, J. (2010). *Giving students effective written feedback.* McGraw-Hill International.

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

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education*, 14(4), 201-206.

Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 128-143). London: Routledge.

Fattah, S. F. E. S. A. (2015). The effectiveness of using WhatsApp messenger as one of mobile learning techniques to develop students' writing skills. *Journal of Education and Practice*, *6*(32), 115-127.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, 11(6), 481-486.

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.

Li, L. & Gao, F. (2020) The effects of social messaging on students' learning and intrinsic motivation in peer assessment. *Journal of Computer Assisted Learning*. DOI: 10.1111/jcal.12409

Güler, Ç. (2017). Use of WhatsApp in higher education: What's up with assessing peers anonymously?. *Journal of Educational Computing Research*, *55*(2), 272-289.

Harrison, K., O'Hara, J., & McNamara, G. (2015). Re-thinking assessment: Self- and peer-assessmentas drivers of self-direction in learning. *Eurasian Journal of Educational Research*, 60, 75–88.

Hashmi, N. (2017). The more the merrier? Understanding the effect of group size on collective intelligence (Doctoral dissertation, Massachusetts Institute of Technology). Retrieved from https://dspace.mit.edu/handle/1721.1/113957

Hung, Y. J., Chen, S. C., & Samuelson, B. L. (2016). Elementary EFL students' practice of peer assessment of oral classroom activities. *TESOL Journal*, *7*(2), 449-468.

Langan, J. (2008). *Ten steps to improving college reading skills* (4th Edition). Beijing: Foreign Language Teaching and Research Press.

Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, 42(4), 645-656.

Li, L. (2018). Using game-based training to improve students' assessment skills and intrinsic motivation in peer assessment. *Innovations in Education and Teaching International*, 56(4), 423-433.

Li, L., & Gao, F. (2016). The effect of peer assessment on project performance of students at different learning levels. *Assessment & Evaluation in Higher Education*, 41(6), 885-900.

Li, L., Liu, X., & Steckelberg, A. L. (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., Liu, X., & Zhou, Y. (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., & Steckelberg, A. L. (2005). Impact of technology-mediated peer assessment on student project quality. *In Proceedings of the Association for Educational Communications and Technology International Conference* (pp. 307-313). Retrieved from http://www.academia.edu/download/30923842/ED499958.pdf#page=317

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). Retrieved from http://www.cjlt.ca/index.php/cjlt/article/view/497/228

Laerd Statistics (2015). One-way MANOVA using SPSS statistics. Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Lin, W. C., & Yang, S. C. (2011). Exploring students' perceptions of integrating Wiki technology and peer feedback into English writing courses. *English Teaching: Practice and Critique*, 10(2), 88-103.

Liu, E. Z.-F., Lin, S. S. J., Chiu, C.-H., & Yuan, S.-M. (2001). Web-Based peer review: the Learner as both adapter and reviewer. *IEEE Transactions on Education*, 44(3), 246-251.

Liu, E. Z., Lin, S. S., & Yuan, S. M. (2002). Alternatives to instructor assessment: A case study of comparing self and peer assessment with instructor assessment under a networked

innovative assessment procedures. *International Journal of Instructional Media*, *29*(4), 395-404.

Lu, Y., Warren, J., Jermaine, C., Chaudhuri, S., & Rixner, S. (2015, May). Grading the graders: Motivating peer graders in a MOOC. In Proceedings of the 24th International Conference on World Wide Web (pp. 680-690). *International World Wide Web Conferences Steering Committee.*

Luo, T., Dani, D. E., & Cheng, L. (2016). Viability of using Twitter to support peer instruction in teacher education. International Journal of Social Media and Interactive Learning Environments, 4(4), 287-304

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1), 48-58.

Mertler, C.A. and Vannatta, R.A. (2005) Advanced and multivariate statistical methods practical application and interpretation. 3rd Edition, Pyrczak Publishing, Los Angeles.

Pascoe, L., Spencer-Smith, M., Giallo, R., Seal, M. L., Georgiou-Karistianis, N., Nosarti, C., ... & Anderson, P. J. (2018). Intrinsic motivation and academic performance in school-age children born extremely preterm: the contribution of working memory. *Learning and Individual Differences*, *64*, 22-32.

Peng, J. C. (2010). Peer assessment in an EFL context: Attitudes and correlations. *In Selected Proceedings of the 2008 Second Language Research Forum, ed. Matthew T. Prior et al* (pp. 89-107).

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.

Rada, R., Acquah, S., Baker, B., & Ramsey, P. (1993). Collaborative learning and the MUCH system. *Computers and Education*, 20(3), 225-233.

Read, A. (2016, April 12). *The biggest Internet phenomenon since the App Store: How messaging Apps are changing Social Media*. Retrieved from https://blog.bufferapp.com/messaging-apps

Rogers, T., & Link, T. (1996). Amazing applications of probability and statistics. Retrieved from http://www.intuitor.com/statistics/SmallGroups.html

Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: a case Study. *Journal of Computer-Based Instruction*, 20(3), 75-80.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and inter-personal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45, 736–750.

Scruggs, T. E., & Mastropieri, M. A. (1998). Tutoring and students with special needs. In K. J. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 165–182) Mahwah, NJ: Lawrence Erlbaum Associates.

Self-Determination Theory. (n.d.). *Intrinsic motivation Inventory* [PDF file]. Retrieve from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

Li, L. & Gao, F. (2020) The effects of social messaging on students' learning and intrinsic motivation in peer assessment. *Journal of Computer Assisted Learning*. DOI: 10.1111/jcal.12409

Shema, H. (2014, April 19). *The birth of modern peer review*. Retrieved from https://blogs.scientificamerican.com/information-culture/the-birth-of-modern-peer-review/

Shehu, V., & Besimi, A. (2017). A proposed model for peer assessment in the digital age: leveraging social medial platforms. *SEEU Review*, 12(2), 173-187.

Shih, R. C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, 27(5), 829-845.

Shen, C. H., & Huang, X. Y. (2006). The Application of Moodle for Web-based peer assessment. *Journal of Educational Media & Library Sciences*, 43(3), 267-284.

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.

Topping, K. J. (2009). Peer assessment. *Theory into practice*, 48(1), 20-27.

Valacich, J., Dennis, A. R., & Nunamaker, J. F. (1992). Group size and anonymity effects on computer-mediated idea generation. Small Group Research, 23(1), 49-73.

Wu, S. Y., Hou, H. T., & Hwang, W. Y. (2012). Exploring students' cognitive dimensions and behavioral patterns during a synchronous peer assessment discussion activity using instant messaging. *Asia-Pacific Education Researcher (De La Salle University Manila)*, 21(3), 442-453.

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.