

The effect of peer assessment on translation performance among EFL students at various achievement levels

L'impatto della valutazione tra pari sulla qualità delle traduzioni in studenti d'inglese come lingua straniera con differenti livelli di rendimento

LAN LI^A, NA ZHAO^{B*}, FEI GAO^C

^A Bowling Green State University, School of Inclusive Teacher Education, Bowling Green, Ohio, USA,
lli@bgsu.edu

^B China Pharmaceutical University, School of Foreign Languages, Jiangning District, Nanjing, China,
zhaona@cpu.edu.cn*

^C Bowling Green State University, Department of Applied Technology, Bowling Green, Ohio, USA,
gaof@bgsu.edu

* corresponding author

HOW TO CITE Li, L., Zhao, N., Gao, F. (2023). The effect of peer assessment on translation performance among EFL students at various achievement levels. *Italian Journal of Educational Technology*. Advance Manuscript Online. Doi: 10.17471/2499-4324/1311

ABSTRACT The purpose of this study was to examine the impact of peer assessment on learning and motivation of students at different achievement (low, average, and high) levels. Forty-one English-as-foreign-language (EFL) students were invited to participate in this study. All students completed an English translation project, then participated in a peer assessment activity to read and comment upon each other's work. Afterward, students revised and resubmitted their own translation projects. Data analysis suggested that the peer assessment activity integrated in this study significantly promoted all students' performance (in terms of increased points between the first version and the second version of students' submissions) on the English translation project, notwithstanding students' achievement levels. Comparisons of students' initial translation projects (before peer assessment) and final translation projects (after peer assessment) revealed that low-achieving students had the greatest degree of improvement, followed by average- and then high-achieving students. Nevertheless, all students reported that they enjoyed the peer assessment activity and put forth comparable effort, as measured by the Enjoyment and Effort subscales. Interpretation of the results and implications were discussed.

Open access, article published by Firenze University Press under CC-BY-4.0 License.

Copyright: © 2023 Author. This is an open access, peer-reviewed article published by Firenze University Press (<https://ijet.itd.cnr.it>) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

KEYWORDS Peer Assessment; Achievement Levels; Motivation; English-as-Foreign-Language (EFL).

SOMMARIO L'obiettivo di questo studio è stato quello di esaminare l'impatto della valutazione tra pari sull'apprendimento e sulla motivazione degli studenti a diversi livelli di rendimento (basso, medio e alto). Quarantuno partecipanti, tutti studenti di inglese come lingua straniera (EFL), sono stati invitati a partecipare a questo studio. Tutti hanno completato un progetto di traduzione in inglese, partecipando in seguito a un'attività di valutazione tra pari, leggendo e commentando il lavoro degli altri. Successivamente, gli studenti hanno rivisto e ripresentato i propri progetti di traduzione. L'analisi dei dati suggerisce che l'attività di valutazione tra pari ha migliorato in modo significativo le prestazioni di tutti gli studenti nella traduzione (in termini di aumento dei voti tra la prima e la seconda versione), indipendentemente dai loro livelli di rendimento. Il confronto tra le traduzioni iniziali degli studenti (prima della valutazione tra pari) e quelle finali (dopo la valutazione tra pari) ha rivelato che gli studenti con risultati bassi hanno avuto il maggior grado di miglioramento, seguiti da quelli con risultati medi e poi da quelli con risultati alti. Ciononostante, tutti gli studenti hanno dichiarato di aver apprezzato l'attività di valutazione tra pari e di essersi impegnati in modo analogo, come misurato dalle scale di gradimento e di impegno. L'interpretazione dei risultati e le implicazioni vengono discusse nel contributo.

PAROLE CHIAVE Valutazione tra Pari; Livelli di Rendimento; Motivazione; Inglese come Lingua Straniera (EFL).

1. LITERATURE REVIEW

1.1. Overview of peer assessment

Peer assessment is a process in which students evaluate the work of their peers and provide feedback using performance criteria (Falchikov, 2007). The use of peer assessment has shown significant growth over the last two decades. The literature reports that peer assessment has been utilized in various disciplines, including teacher education (Li, 2011; Lynch, McNamara, & Seery, 2012), physics (Handayani & Genisa, 2019), computer science (Wang, Li, Feng, Jiang, & Liu, 2012), art education (Liu, Wan, Tu, Chen, & Wang, 2021), medicine education (Violato & Lockyer, 2006), engineering (Hersam, Luna, & Light, 2004), foreign language acquisition (Cheng & Warren, 2005; Hung, 2018), biology (Orsmond, Merry, & Reiling, 1996), and business (Brutus, Donia, & Ronen, 2013; Tiew, 2010). As a learning tool, peer assessment has demonstrated positive effects on student academic achievement, cognitive growth, and other learning or developmental outcomes (Li & Gao, 2016). After an extensive review of peer assessment studies in higher education, Topping (1998) reaches a conclusion that peer assessment may “*yield gains in the cognitive, social, affective, transferable skill, and systemic domains that are at least as good as those from staff assessment*” (p. 269). A more recent meta-analysis of 58 peer assessment studies (Li, Xiong, Hunter, Guo, & Tywoniw, 2020) confirms that peer assessment helps students improve their performance. The researchers further report that computer-assisted peer assessment generates greater learning gains than paper-based peer assessment. The literature suggests that these “learning gains” may include:

- Enhancing students' autonomy, motivation (Brown, 2004; Hiltz & Wellman, 1997; Pope, 2001), course attendance (Deslauriers, Schelew, & Wieman, 2011), and engagement (Handayani & Genisa, 2019; Lucas, 2009).
- Promoting student responsibility (Somervell, 1993).
- Fostering students' critical thinking skills (Zhan, 2021).
- Encouraging interpersonal skills among students (Dochy, Segers, & Sluijsmans, 1999; Patri, 2002).
- Deepening students' understanding of assessment criteria and quality performance (Dochy et al., 1999).
- Improving quality of work (Li & Steckelberg, 2005; Liu et al., 2021; Pope, 2001) and learners' conceptual understanding (Duncan, 2005).
- Supporting collaborative online learning (Zou, Xie & Wang, 2021).

Despite its growing popularity and documented potential, peer assessment is not a strategy without drawbacks. Some of these pitfalls include peer pressure and student assessment skills (Li & Gao, 2016). Take peer pressure as an example, Robinson states (1999, p. 96) that when students assess each other, "*marking could be easily affected by friendship, cheating, ego or low self-esteem.*" This problem is especially prominent in paper-based peer assessment when it is difficult to protect the anonymity of assessors and assessees. The other issue relates to students' ability to conduct valid and valuable assessment of peers' work. The traditional role of a student is a learner who is usually being assessed and rarely assesses others. Switching students' roles from assessee to assessor is generally difficult and may require assistance, such as assessment training, to get students prepared (Li & Gao, 2016).

1.2. Peer assessment and student achievement levels

It remains unclear what makes effective peer assessment (Li & Grion, 2019; Van Zundert, Sluijsmans, & Van Merriënboer, 2010). In the handful of studies that lend insight into the "how students learn" aspect of peer assessment, Davies (2000) noted that the peer assessment activity in his study had a disproportionate impact on students. Comparing students' test scores before and after the peer assessment process showed that students at higher achievement levels benefited the least from the process (by gaining the smallest increase from their pre- to post-report scores), while students at lower achievement levels benefited the most (by gaining the biggest increase from their pre- to post-report scores). This captivating but incidental observation was supported, to some extent, by two later studies (Li, Steckelberg, & Srinivasan, 2008; Li, 2011). One study (Li et al., 2008) explored student perceptions toward peer assessment. While students' attitudes were generally positive toward peer assessment, one high-achieving student stated that, "*sometimes peer assessment isn't helpful if you already did a good job*" (p. 143). The other study (Li, 2011), with a descriptive and qualitative approach, examined the impact of peer assessment on the work of students at different achievement levels. Their discovery was in line with the studies discussed above that peer assessment seemed to benefit students in the early academic development stages more than students at more advanced achievement levels.

As intriguing as these observations were, they were either incidental, or from qualitative or descriptive analyses, which bear the limitation that the knowledge gained from these studies may not be generalized beyond the specific samples of these studies. To circumvent the limitation, Li and Gao (2016) took the investigation one step further

and used a two-way factorial design to investigate how peer assessment interplayed with student achievement levels to impact student learning. One hundred and thirty undergraduate students were assigned into either an experimental group or a control group working on a technology-integrated lesson plan. While students in the control group used a traditional discussion approach to improve their lesson plans, students in the experimental group reviewed peers' projects and provided feedback to help each other improve their own projects. Based on the scores of the students' first draft lesson plans, students in each group (control or experimental) were divided into three levels: low, average, and high achieving. The comparisons of the students' final lesson plan scores in both groups suggested that the peer assessment group outperformed the control group on their lesson plan project scores. Data analysis also indicated a joint influence of peer assessment and student achievement levels on the quality of their lesson plan projects. More specifically, the findings were in agreement with other studies by showing that low- and average-achieving students reported a bigger learning gain than high-achieving students.

While encouraged by the confirmed benefits of peer assessment on student learning, the researchers (Li & Gao, 2016) noted a possible measurement limitation in their study: the ceiling effect. They speculated that the ceiling effect may explain the findings of the limited learning gain of high-achieving students in their study. Specifically, the researchers speculated that that *"...high achievers' scores were close to the maximum score on the lesson plan project, which may have created a ceiling effect and may have decreased the likelihood that the rubric had accurately measured learning gains of high achievers. In other words, high achievers may simply have not had enough room to demonstrate their improvement."* According to Cramer and Hotwitt (2004), *"failure to recognize the possibility that there is a ceiling effect may lead to the mistaken conclusion that the independent variable has no effect"* (p. 21). The current study aims to address the possible ceiling effect issue by increasing the difficulty level of the project for students to complete in a peer assessment activity, so most students will not be able to score near or close to the maximum score. To attain this goal, the researchers purposefully identified an English translation activity in which students could employ peer assessment to improve their scores.

1.3. Peer assessment in translation instruction

Researchers and educators in the translation field have long noticed that the traditional teacher-centered translation classes do not motivate or inspire students (Colina, 2003; Davis, 2005; Kiraly, 1995; Kiraly, 2000) and have been exploring ways to actively engage students in learning. Although peer assessment has been well studied in second or foreign language writing and speaking instruction (Rollinson, 2005; Chien, Hwang, & Jong, 2020; Shen, Bai, & Xue, 2020), only a handful of researchers examined the effects of peer assessment in translation instruction (Wang & Han, 2013; Huynh & Nguyen, 2019; Lin, Song, Guo, & Wang, 2021). These studies revealed that students viewed peer assessment as positive (Wang & Han, 2013; Huynh & Nguyen, 2019), were motivated during the process (Huynh & Nguyen, 2019), and acknowledged an improvement in their translation skills (Wang & Han, 2013; Huynh & Nguyen, 2019). Studies also suggested that peer-assessment-based translation activities engaged students in collaborative learning (Insai & Poonlarp, 2017; Lin, Song, Guo, & Wang, 2021), offered students different perspectives on solving problems in their translation (Wang & Han, 2013; Huynh & Nguyen, 2019), empowered students in their learning, and fostered critical thinking (Wang & Han, 2013). The findings suggest that there is value in integrating peer assessment into translation instruction and that students with various

achievement levels are impacted in different ways, as discussed above. However, there are still questions left unexplored. In particular, it is unclear whether peer assessment has an equal impact on students' performance regardless of students' achievement levels, or what instructors should do to accommodate students with different achievement levels in peer-assessment-based translation activities.

1.4. Purpose statement and research questions

The current study will contribute to the existing research literature in two ways. First, the study will add to the peer assessment research by addressing the ceiling effect in the previous study (Li & Gao, 2016). The researchers believed that the translation project was at an appropriate difficulty level, and it would be unlikely for most students to achieve very high scores in their initial drafts before peer assessment. Second, the study will add to the limited research on how to improve translation instruction through peer assessment. In light of the literature, the current study seeks to answer three research questions:

- 1) Are there statistically significant differences in learning gain in terms of increased grades (from the initial version to the final version of a student English translation project) in a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?
- 2) Are there statistically significant differences in student self-reported enjoyment during a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?
- 3) Are there statistically significant differences in student self-reported effort put forth during a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?

2. METHODOLOGY

2.1. Participants and settings

Forty-one undergraduate students (6 males and 35 females) at a university in southeast China were invited to participate in this study. All participants were English majors and were enrolled in two sections of an English translation class. The two class sections were taught by the same instructor, using the same curriculum. Of those reported, the students' ages were between 19 and 21 with a mean of 20.6.

This study was approved by the IRB (Institutional Review Board) at Bowling Green State University (ID: 978720-3). Students were invited to attend an information session to learn about the research project and understand their rights as human subjects. Only students who signed the consent form were invited to participate in this study.

2.2. Procedure

In the three-week-long lesson plan module, all students completed an assignment of translating a passage of about 450 Chinese characters into English. The translation task was intentionally designed to be moderately challenging to eliminate the possibility of too many low or high scores in students' initial drafts. The researchers anticipated that this approach would mitigate the potential ceiling effect. After students completed their first draft translation,

they followed the peer assessment procedure. In week 1, students submitted their work to an online document collaboration site, called Shimo (<https://shimo.im/>) (initial submission). A cloud-based productivity tool called Shimo Document was deliberately selected to enhance and facilitate the peer assessment process. Shimo offers a range of features to accelerate collaboration and online document editing. Similar to Google Docs, this platform supports multi-end synchronization and allows users to create, edit, and share documents in real-time.

All students received training on how to evaluate peer projects and provide feedback. Students were instructed that the project rubric should be considered when they provide feedback to peers. Feedback should be backed up by the rubric. In week 2, each student was randomly assigned to review and comment on five or six translation submissions from peers within Shimo. Each student posted at least two comments to each assigned work and also responded to their classmates' comments on their own work. At this stage, the students were instructed that they did not have to blindly follow suggestions or comments from peers. The students understood, from the previous training, that the quality of peer feedback varied. They were strongly encouraged to evaluate the quality of peer feedback with a reference to the rubric and only adopt suggestions that they deemed correct. In week 3, students revised their translation work based on their classmates' comments and their reflections on their classmates' work (final submission). Students then completed a survey that included statements to measure their enjoyment and effort put forth in the peer assessment activity.

2.3. Data collection and analysis

2.3.1. Student learning

The researchers looked at learning gain in terms of increased grades from the first-round project grade to the second-round project grade. The instructor graded all the submissions using a translation evaluation rubric (see Table 1), and a score ranging from 0 to 10 was assigned to each student's work. Decimal scores were allowed for grading. For example, a student's essay score could be 7.6 out of 10. To ensure the reliability of grading, the course instructor and a second rater worked together to grade 25% of randomly selected assignments. They first reviewed and discussed the rubrics to reach a common understanding of grading criteria. Next, they independently graded ten randomly selected assignments based on the rubrics. Then, they compared their gradings and resolved the disagreements through discussion. To determine the agreement between two raters, the researchers computed the percent of exact or adjacent agreement (within 1 point) between the two grading sets. The percentage of exact agreement was 45.00%, while the percentage of adjacent agreement was 95.00%, which suggested a good level of consensus (Jonsson & Svingby, 2007; Stemler, 2004).

2.3.2. Student motivation

To measure student enjoyment and effort levels in the peer assessment activity, this study used two subscales from the Intrinsic Motivation Inventory (IMI) (Ryan, 1982). The IMI has been tested as reliable and valid (McAuley, Duncan, & Tammen, 1989), and has been extensively adopted to gauge student motivation, especially in technology-enabled learning environments (Bertacchini, Bilotta, Pantano, & Tavernise, 2012). Two subscales,

Enjoyment and Effort, were selected in this study to measure students' perceptions in the peer assessment activity. Both subscales are rated on a 7-point Likert scale ranging from 1 (not at all true) to 7 (very true). The Enjoyment subscale consists of six items measuring students' self-reported intrinsic motivation. Example items include “*I thought this activity was quite enjoyable,*” and “*This activity was fun to do*”. The Effort subscale consists of five items that gauge the effort students put forth to complete the peer assessment project. Items include “*I put a lot of effort into this,*” and “*I didn't try very hard to do well at this activity*”. Reverse coding was applied for negatively framed statements.

Categories	Criteria
Excellent 9-10 points	The translation faithfully reflects all of the original passage with only 1 or 2 minor errors in vocabulary, syntax, punctuation or spelling. The translation is elegant (appropriate choice of words, a variety in sentence patterns).
Good 7-8 points	The translation reflects almost all the original passage with relatively few significant errors of vocabulary, syntax, spelling or punctuation. The translation is readable (generally clear, smooth and cohesive).
Acceptable 5-6 points	The translation adequately reflects most of the original passage with occasional errors of vocabulary, syntax, spelling or punctuation. The translation is, for the most part, readable.
Inadequate 3-4 points	The translation only reflects about half of the original passage with frequent errors of vocabulary, syntax, spelling or punctuation. The translation is, in some parts, unreadable.
Poor 1-2 points	The translation reflects less than half of the original passage. Almost all sentences contain errors of vocabulary, syntax, spelling or punctuation. The translation is, for the most part, unreadable.

Table 1. Student translation evaluation rubric.

3. Results

3.1. Degree of Improvement in revised translation project

Research Question 1: “*Are there statistically significant differences in Improvement score (change of score from the initial version to the final version of a student English translation project) after a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?*”

A one-way ANOVA was conducted to determine if the degree of improvement (Improvement score) from the initial versions of the students' translation project to the final versions was different for groups at different achievement levels. Participants were assigned to three levels based on the scores of their initial versions of the project and their TEM-4 (Test for English Majors, Band-4) scores: low achieving (n = 15), average achieving (n

= 14), and high achieving (n = 12). TEM-4 is a required national standardized test to measure English proficiency of English major undergraduate students in China. The test was taken at the end of the sophomore year, before students participated in this study. TEM-4 is generally considered a reliable and valid test of student English proficiency (Jin & Fan, 2011). A test validation study (TEM Test Center, 1997, p. 63) suggested that “*the TEM tests are reasonably reliable and valid tests that are set at an appropriate (difficult) level as defined in the test specification*”.

A few assumption tests were conducted to determine if it would be appropriate to use one-way ANOVA to answer the first research question. Boxplot suggested no outliers in the data. Data were normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). Levene's test of homogeneity of variances ($p = .161$) indicated that there was homogeneity of variances.

Data are presented as mean \pm standard deviation. The Improvement score that measures the difference between the initial submission score and the final submission score of the student English translation project decreased from the low-achieving (1.90 ± 0.66) to the average-achieving (1.71 ± 0.51), and to the high-achieving (1.17 ± 0.44) level, in that order. As Table 2 shows, the Improvement score was statistically significantly different for different achievement levels, $F(2, 38) = 6.16$, $p < .005$. Tukey post hoc analysis revealed that there was a decrease in the Improvement score from the low-achieving group to the high-achieving group, with a decrease of 0.73 ± 0.21 [mean \pm standard error], which was statistically significant ($p = .004$). In addition, there was a decrease in the Improvement score from the average-achieving group to the high-achieving group with a decrease of 0.55 ± 0.22 [mean \pm standard error], which was also statistically significant ($p = .042$). However, the group difference between low achieving and average achieving was not statistically significant.

Source	df	SS	MS	F	p
Between Groups	2	3.77	1.88	6.16	.005
Within Groups	38	11.62	0.31		
Total	40	15.39			

Note: df = degree of freedom; SS = Sum of Squares; MS = Mean Square

Table 2. One-Way Analysis of Variance of Improvement Score by Student Achievement Levels.

3.2. Motivation Scales

A one-way multivariate analysis of variance (MANOVA) was initially considered to determine the effect of peer assessment on student attitudes. Two measures of student attitudes were assessed: self-reported Enjoyment score and Effort score. Students were from three achievement levels: low-, average-, and high-achieving levels.

The Pearson Correlation showed negligible correlation ($r = 0.178$, $p = .265$) between the two dependent variables: Enjoyment score and Effort score. Although the Pearson Correlation showed no evidence of multicollinearity, the correlation suggested that the MANOVA test was not suitable to determine the effects of peer assessment on the two dependent variables, as the dependent variables should be moderately correlated in MANOVA (Laerd Statistics, 2015; Salkind, 2010). Salkind further states, “*If there is no correlation at all, MANOVA offers no*

improvement over an analysis of variance (ANOVA).” Therefore, the researchers decided to run two separate ANOVAs to determine the impact of peer assessment on students’ Enjoyment and Effort scores.

Research Question 2: “*Are there statistically significant differences in Enjoyment score after a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?*”

A one-way ANOVA was conducted to determine if the self-reported Enjoyment score was different for different levels: low achieving (n = 15), average achieving (n = 14), and high achieving (n = 12). Assumption tests showed that there were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and Levene’s test of homogeneity of variances ($p = .952$) indicated homogeneity of variances. Data is presented as mean \pm standard deviation. As Table 3 shows, the Enjoyment score varied from low achieving (31.53 ± 5.17) to average achieving (30.79 ± 5.42), to high achieving (32.33 ± 5.14), in that order, but the differences between these achievement levels were not statistically significant, $F(2, 38) = 0.281$, $p = .757$.

	Level	M	SD
Enjoyment	Low Achieving	31.53	5.17
	Average Achieving	30.79	5.42
	High Achieving	32.33	5.14
Effort	Low Achieving	26.80	3.73
	Average Achieving	27.43	3.35
	High Achieving	29.50	4.85

Note. M = Mean; SD = Standard Deviation

Table 3. Means and Standard Deviations for Low-Achieving, Average-Achieving, and High-Achieving Levels on the Students’ Enjoyment and Effort Scores.

Research Question 3: “*Are there statistically significant differences in Effort score after a peer assessment activity between students at low-, average-, and high-achieving levels in a translation class?*”

A one-way ANOVA was conducted to determine if the self-reported Effort score was different for different levels: low achieving (n = 15), average achieving (n = 14), and high achieving (n = 12). Assumption tests showed that there were three outliers, as assessed by boxplot. The researchers decided to include the outliers in the analysis, as the ANOVA test was run twice with and without the outliers and both tests yielded the same results. Data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and Levene’s test of homogeneity of variances ($p = .608$) indicated that there was homogeneity of variances. Data is presented as mean \pm standard deviation. As Table 3 shows, the Effort score increased from low achieving (26.80 ± 3.73) to average achieving (27.43 ± 3.35), to high achieving (29.50 ± 4.85), in that order, but the difference between these achievement levels was not statistically significant, $F(2, 38) = 1.638$, $p = .208$.

4. DISCUSSION

Data analysis revealed that peer assessment had a positive effect on the performance of all students regardless of their achievement levels, as all students' final translation scores improved from their initial scores after peer review. Among them, the performance of students at the low-achieving level improved the most, followed by students at the average-achieving level, and then students at the high-achieving level. Data also indicated that the score change between the low-achieving and high-achieving groups, and between the average-achieving and high-achieving groups was significant. However, although both Enjoyment and Effort scores generated by the high-achieving group were higher in numerical value than those from the low- and average-achieving groups, comparisons of both scores between the three levels showed no significant difference.

The findings of this study support the previous peer assessment research (Davies, 2000; Li & Gao, 2016) in that peer assessment may have unequal influence on student learning. The current study examined the relationship between peer assessment and student achievement levels from a different angle by comparing the degree of improvement between three student achievement levels: low achieving, average achieving, and high achieving. While all students, regardless of their achievement levels, improved their projects more or less after peer assessment, the comparisons between students' first-round and second-round submitted projects confirmed that the low-achieving students had the biggest improvement, followed by average-achieving students and then high-achieving students.

Interestingly, although the leveled effect of peer assessment on students' learning was supported by the statistical findings of this study, the comparison of mean scores of students' Enjoyment and Effort ratings did not show a significant difference between groups. These findings (significant difference on the change scores between groups and non-significant difference on Enjoyment and Effort scores between groups) seem conflicting but actually are not surprising. First, all three groups — low, average and high achieving — improved their performance on their final project submissions after peer assessment. Even though the degree of improvement as represented by the change score (between the initial and final project submissions) for the high-achieving group was not as high as that of the low-achieving and average-achieving groups, a further look at the high-achieving group data indicated a significant difference between its first- and second-round project submissions ($t_{11} = 9.100, p < 0.001$). In other words, students in the high-achieving group significantly improved their translation projects after participating in the peer assessment activity. On average, the second-round submissions were assessed 1.17 points higher than their initial submissions (95% CI [.89, 1.45]). This finding suggests a possible explanation for why the ratings of Enjoyment and Effort statements by the high-achieving students were similar to those of the other two groups. Since all three groups performed significantly better on their projects after peer assessment, it is not surprising that their intrinsic motivation rating (as represented by the Enjoyment score) showed no significant difference. The results also suggested that all students, regardless of their achievement levels, put forth comparable effort (as represented by the Effort score) into completing the peer assessment activity. It would be interesting for future studies to analyze if students demonstrate different patterns of participation and interaction across different achievement levels in a peer assessment activity.

5. CONCLUSION

This study examined how peer assessment may impact learning of students at different achievement levels in translation instruction. The findings of the study are in line with the results of the previous study on peer assessment (Li & Gao, 2016) and presented three relevant facts:

- 1) The peer assessment activity used in this study was effective in enhancing students' performance on an English translation project, regardless of students' achievement levels.
- 2) The peer assessment activity had unequal impact on learning of low achievers, average achievers and high achievers.
- 3) Comparisons of students' first-round project assessment (before peer assessment) and second-round project assessment (after peer assessment) showed that low achievers had the greatest degree of improvement, followed by average and then high achievers.

The significance of the study lies in the fact that this is an innovative study that investigated the impact of peer assessment on students at various achievement levels in translation instruction. Although the potentials of peer assessment are widely studied and well documented, it is still not fully understood what makes effective peer assessment (Li & Grion, 2019; Van Zundert et al., 2010). This paper is the second in a series on the "how peer assessment works" inquiries. The initial paper (Li & Gao, 2016) discovered an interaction between peer assessment and student achievement levels and suggested that low achievers may have the biggest learning gain from peer assessment, while high achievers have the least. This paper was specifically designed

- 1) to address one possible limitation in the initial paper, ceiling effects, and
- 2) to examine the impact of peer assessment on students' performance from a different angle by comparing the degree of performance improvement between the three achievement groups: low, average, and high.

The findings of this study confirmed that peer assessment had unequal impact on students at various achievement levels. What is more, the current papers indicated that high achievers also had significant learning gain from peer assessment, even though their learning gain may not be as much as that of the low and average achievers.

The finding of peer assessment significantly improving the learning of high achievers in this study seems contradictory to the finding of our previous study (Li & Gao, 2016) that suggested peer assessment benefits low or average achievers but not necessarily high achievers. However, the seemingly conflicting findings of the two studies may be explained, to some extent, by the ceiling effect theory. As discussed in the Literature Review section, the researchers (Li & Gao, 2016) speculated that "...high achievers' scores were close to the maximum score on the lesson plan project, which may have created a ceiling effect and may have decreased the likelihood that the rubric had accurately measured learning gains of high achievers." In other words, the researchers (2016) suspected that the lesson plan assignment was insufficiently difficult to measure high achievers' true knowledge or ability, as most of them made it close to the maximum score in their first submissions. The speculation of a ceiling effect existing in the previous article (2016) seemed to be confirmed by the findings of this study. An English translation project was purposefully chosen as the measure in this study, as the researchers believed that the project was at an appropriate difficulty level and it would be unlikely for most students to achieve very high scores in their initial drafts before peer assessment. It turned out to be an effective measure that lessened the ceiling

effect and more accurately reflected the achievement of learners including high achievers. These two studies show that student achievement level is an important factor to consider in peer review activities. This conclusion opens up potential for researchers and educators to experiment with different ways to group students based on their achievement levels. For example, the instructor may consider intentionally grouping students at different achieving levels in the same group.

The findings of this study have important practical implications for educators and researchers who are interested in peer assessment in translation instruction or other fields. This study indicated that peer assessment is an appropriate instructional approach to benefit students at all achievement levels in the translation field. However, assessment criteria of translation assignments that are used to measure student learning in peer assessment should have appropriate difficulty levels that can accurately gauge student performance. If assessment tools are not sufficiently sensitive or reliable for distinguishing a difference in quality of translation work, students' immediate learning gains may not readily manifest, even though they may exist. While the details of the translation task are specific to language-learning students, the general principle of selecting activities with suitable difficulty levels to appropriately measure student performance applies across disciplines.

Given the cultural context of Chinese students, it's worth considering how cultural norms around giving and receiving feedback might influence their peer assessment process. For example, in some Asian cultures, young people endorse interpersonal harmony, hierarchical relationships and traditional conservatism (Zhang, Lin, Nonaka, & Beom, 2005). They may prefer providing and receiving indirect or subtle feedback to maintain harmony, cohesion or save face. This could potentially impact the nature and tone of the feedback provided by the students in peer assessment. Researchers and educators should consider these cultural nuances when interpreting the results of the current study. Future studies are warranted to replicate the study findings with subjects in other culture contexts.

In the past a few decades, technology has shaped and reshaped the education sector including the peer assessment field. In this study, Shimo Document was chosen as the collaboration platform to facilitate the peer assessment process. This productivity suite that is equipped with different features such as chat, documents, spreadsheets in a simple interface stood out as a robust solution to encourage interactions between students. In addition, Shimo's cloud-based nature facilitates both real-time and asynchronous collaborations and allows students to engage in the peer assessment process at their own time and pace. This anytime and anywhere feature can be particularly beneficial in accommodating students' diverse learning needs and various schedules.

While the study discloses some interesting findings, the researchers would like to acknowledge two possible limitations on the generalization of these findings. First, this study employed a specific group of EFL students who followed a specific peer assessment model to improve an English translation project at a Chinese university. The findings of the study may not be generalized to other peer assessment studies or other populations. The researchers call for future studies to determine if the study could be replicated and similar findings generated in other populations or with different peer assessment models. Second, this study used a fairly small sample that decreases statistical power and the flexibility of the effect size. Future studies are warranted to confirm the findings with bigger samples. Future studies should also employ a control group that does not follow the peer assessment procedure and should measure the enjoyment and effort scores before and after peer assessment. It would be interesting to see how these scores may differ between the treatment group and the control group, and before and

after peer assessment activity. Designs that employ other research approaches, such as qualitative and mixed-methods, should also be considered in future studies so data triangulation can be used to help better understand student experiences when peer assessment is used in translation instruction.

References

- Bertacchini, F., Bilotta, E., Pantano, P., & Tavernise, A. (2012). Motivating the learning of science topics in secondary school: A constructivist edutainment setting for studying chaos. *Computers & Education*, 59(4), 1377–1386. <https://doi.org/10.1016/j.compedu.2012.05.001>
- Brown, D. (2004). *Language assessment: Principles and classroom practice*. New York, NY: Longman.
- Brutus, S., Donia, M. B., & Ronen, S. (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. Retrieved from <https://www.jstor.org/stable/23412389>
- Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education*, 146. <https://doi.org/10.1016/j.compedu.2019.103751>
- Colina, S. (2003). *Translation teaching, from research to the classroom: A handbook for teachers*. New York, NY: McGraw-Hill.
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93-121. <https://doi.org/10.1191/0265532205lt29>
- Cramer, D., & Howitt, D. (2004). *The SAGE dictionary of statistics* (Vols. 1-0). London, UK: SAGE. doi: 10.4135/9780857020123
- Davies, M. G. (2005). Minding the process, improving the product: Alternatives to traditional translator training. In M. Tennent (Ed.). *Training for the new millennium: Pedagogies for translation and interpreting* (Benjamins Translation Library 60, pp. 67–82). Amsterdam, NL: John Benjamins.
- Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International*, 37(4), 346-354. <https://doi.org/10.1080/135580000750052955>
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332, 862-864. <https://doi.org/10.1126/science.1201783>
- Dochy, F., Segers, M., & Sluijsmans, D. M. A. (1999). The use of self-, peer-, and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>

- Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems*. San Francisco, CA: Pearson/Addison-Wesley.
- 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). New York, NY: Routledge.
- Handayani, R. A. D., & Genisa, M. U. (2019). Empowering Physics Students' Performance in a Group Discussion through two types of peer assessment. *International Journal of Instruction*, 12(1), 655-668. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1201315.pdf>
- Hersam, M. C., Luna, M., & Light, G. (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., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, 40(2), 44-49.
- Hung, Y. J. (2018). Group peer assessment of oral English performance in a Taiwanese elementary school. *Studies in Educational Evaluation*, 59, 19-28. <https://doi.org/10.1016/j.stueduc.2018.02.001>
- Huynh, T. L., & Nguyen, U. N. T. (2019). Peer assessment in a blended translation course: Students' perceptions, motivation and their self-perceived translational skill development. *International Journal of Applied Linguistics and English Literature*, 8(5), 52-60. <https://doi.org/10.7575/aiac.ijalel.v.8n.5p.52>
- Insai, S., & Poonlarp, T. (2017). More heads are better than one: Peer editing in a translation classroom of EFL learners. *PASAA: Journal of Language Teaching and Learning in Thailand*, 54, 82-107. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1171212.pdf>
- Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, 28(4), 589-596. <https://doi.org/10.1177/0265532211414>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kiraly, D. (1995). *Pathways to translation: Pedagogy and process*. Kent, OH: The Kent State University Press.
- Kiraly, D. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. London, UK: Routledge.
- Laerd Statistics. (2015). *One-way MANOVA in SPSS Statistics. Statistical tutorials and software guides*. Retrieved from <https://statistics.laerd.com/spss-tutorials/one-way-manova-using-spss-statistics.php>

- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193-211.
<https://doi.org/10.1080/02602938.2019.1620679>
- Li, L. (2011). How do students of diverse achievement levels benefit from peer assessment? *International Journal for Scholarship of Teaching and Learning*, 5(2). <https://doi.org/10.20429/ijstl.2011.050214>
- 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.
<https://doi.org/10.1080/02602938.2015.1048185>
- Li, L., & Grion, V. (2019). The power of giving feedback and receiving feedback in peer assessment. *All Ireland Journal of Higher Education*, 11(2). Retrieved from <https://ojs.aishe.org/index.php/aishe-j/article/view/413>
- Li, L., & Steckelberg, A. L. (2005). Impact of technology-mediated peer assessment on student project quality. In Simonson, M. & Crawford, M. (Eds.), *Proceedings of the Association for Educational Communications and Technology International Conference 2005* (pp. 307-313). Orlando, FL.
- Li, L., Steckelberg, A. L., & Srinivasan, S. (2008). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology*, 34(2), 133-148.
<https://doi.org/10.21432/T21C7R>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193-211.
<https://doi.org/10.1080/02602938.2019.1620679>
- Lin, Z., Song, X., Guo, J., & Wang, F. (2021). Peer feedback in translation training: A quasi-experiment in an advanced Chinese–English translation course. *Frontiers in Psychology*, 12. doi: 10.3389/fpsyg.2021.631898
- Liu, C., Wan, P., Tu, Y. F., Chen, K., & Wang, Y. (2021). A WSQ-based mobile peer assessment approach to enhancing university students' vocal music skills and learning perceptions. *Australasian Journal of Educational Technology*, 37(6), 1-17. <https://doi.org/10.14742/ajet.6832>
- Lucas, A. (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus*, 19(3), 219–231. <https://doi.org/10.1080/10511970701643970>
- Lynch, R., McNamara, P. M., & Seery, N. (2012). Promoting deep learning in a teacher education programme through self-and peer-assessment and feedback. *European Journal of Teacher Education*, 35(2), 179-197.
<https://doi.org/10.1080/02619768.2011.643396>

- 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. doi:10.1080/02701367.1989.10607413
- Orsmond, P., Merry, S. & Reiling, K. (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.
- Robinson, J. (1999). Computer-assisted peer review. In S. Brown, J. Bull, & P. Race (Eds.), *Computer-assisted assessment in higher education* (pp. 95-102). London, UK: Kogan Page
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23-30.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.
- Salkind, N. J. (2010). *Encyclopedia of research design* (Vols. 1-3). Thousand Oaks, CA: SAGE.
<https://doi.org/10.4135/9781412961288>
- Shen, B., Bai, B., & Xue, W. (2020). The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class. *Studies in Educational Evaluation*, 64.
<https://doi.org/10.1016/j.stueduc.2019.100821>
- 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.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, 9(4), 1–11.
- TEM Test Center (1997). *The Test for English Majors (TEM) validation study*. Shanghai, CN: Shanghai Foreign Language Education Press.
- Tiew, F. (2010). Business students' views of peer assessment on class participation. *International Education Studies*, 3(3), 126-131.
- van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.

- Violato, C., & Lockyer, J. (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.
- Wang, K., & Han, C. (2013). Accomplishment in the multitude of counsellors: Peer feedback in translation training. *Translation & Interpreting*, 5(2), 62-75. <https://doi.org/doi: 10.12807/ti.105202.2013.a05>
- Wang, Y., Li, H., Feng, Y., Jiang, Y., & Liu, Y. (2012). Assessment of programming language learning based on peer code review model: Implementation and experience report. *Computers & Education*, 59(2), 412-422. <https://doi.org/10.1016/j.compedu.2012.01.007>
- Zhan, Y. (2021). What matters in design? Cultivating undergraduates' critical thinking through online peer assessment in a Confucian heritage context. *Assessment & Evaluation in Higher Education*, 46(4), 615-630. <https://doi.org/10.1080/02602938.2020.1804826>
- Zhang, Y., Lin, M., Nonaka, A., & Beom, K. (2005). Harmony, hierarchy and conservatism: A cross-cultural comparison of Confucian values in China, Korea, Japan, and Taiwan. *Communication Research Reports*, 22(2), 107-115. <https://doi.org/10.1080/00036810500130539>
- Zou, D., Xie, H., & Wang, F. L. (2021). Peer-assessment enhanced collaborative learning in a virtual learning environment. *International Conference on Blended Learning* (pp. 132-141). Cham, CH: Springer. https://doi.org/10.1007/978-3-030-80504-3_11