

# Investigating the Validity of Different Peer Groupings in the Assessment of English Writings

Zhou Bi<sup>1</sup>, Tan Mengyu<sup>1</sup>, Yuan Diying<sup>1</sup> & Pan Liandi<sup>1</sup>

<sup>1</sup> Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Zhou Bi, Guangdong University of Foreign Studies, No.2 Baiyun Avenue, Guangzhou City, China.

Received: July 11, 2019

Accepted: September 13, 2019

Online Published: November 29, 2019

doi:10.5539/ies.v12n12p55

URL: <https://doi.org/10.5539/ies.v12n12p55>

## Abstract

Peer assessment is an indispensable part in classroom assessment and it serves as a very useful way of promoting learning. However, different ways of peer grouping may influence the validity of peer assessment. This study analyzes the quality and quantity of feedback, adoption rate of feedback as well as scores of students' original drafts and the revised versions. It finds that all ways of grouping can promote learning but the degree of validity varies among groups. Besides, accuracy and adoption rate are high in students' feedback, which means peer feedback is effective to a great extent. Among all the ways of grouping, homogeneous grouping i.e. pairing students with the same or similar language proficiency level can archive more significant promotion in learning. In general, students hold a positive view towards the validity of peer assessment.

**Keywords:** validity, peer assessment, peer grouping, peer feedback, learning

## 1. Introduction

### 1.1 Research Background

Recently, the Ministry of Education of China is actively promoting the establishment of China's Standards of English Language Ability (CSE) and a foreign language proficiency assessment system, which aims at facilitating both formative assessment and summative assessment.

Writing, as one of four basic skills for ESL/EFL students, has become one of the most important research field facilitating the establishment of assessment system. Many scholars probe into what kind of assessment methods would be beneficial for improving students' writing competence. In previous studies, the assessment methods applied to writing mainly had three strands, namely teacher assessment, self-assessment and peer assessment. Among these assessment, teacher assessment is regarded as the most conventional way to give feedback on students' writing, but it features some weaknesses - long time spent on revising students' writings, no immediate feedback from teachers and response from students, difficulty in expressing their ideas under the authority of teachers and so on. Self-assessment is regarded as the most flexible method, but it seems hard for students to revise their own writings due to the limit of their language proficiency.

As a common method of formative assessment, peer assessment has attracted wide attention at home and abroad. Many studies have focused on exploring the application of peer assessment and demonstrating the benefits that learners can get from peer assessment (Li, 2014). However, there also exist some drawbacks on applying peer feedback on writing. Some students are poor at writing and unable to give accurate and sufficient feedback to their peers. Besides, compared with teacher feedback, peer feedback is questioned in terms of the effectiveness. Therefore, there is a need to investigate an optimal model of peer feedback. Few studies, however, have explored which types of peer assessment grouping are more likely to promote learning. Therefore, this study aims to explore the effectiveness of different groupings in peer assessment so as to construct an optimal peer assessment model.

### 1.2 Literature Review

Peer assessment is a way of collaborative learning, in which two or more students form a learning group to help each other with their learning (Li & Ke, 2013). When conducting peer assessment, students give feedback and suggestions on their peers' work and discuss the difficulty they encounter during the process of the learning. Peer assessment is theoretically underpinned by the Zone of Proximal Development, which proposes that there is a "distance between the actual developmental level as determined by independent problem-solving ability and the

level of potential development as determined through problem-solving ability under adults' guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Therefore, within learners' potential area of development, learning can happen with peers' collaboration or scaffolding of more capable guiders (Ma, 2005; Ma, 2012). It reviews the effectiveness of peer assessment for learning and it also suggests that peer collaboration can promote students' learning enthusiasm.

Research on formative assessment has been few in China. Most of the studies discuss the construction of a theoretical framework and the feasibility of formative evaluation (Li & Zeng, 2008; Wang & Zhen, 2008; Li, 2012; Yang & Wen, 2013; Li, 2014). Only in the past decade, empirical research on formative assessment has increased, mostly on its application in college English courses (Wen, 2011; Yang & Wen, 2014a; Yang & Wen, 2014b; Li, 2010; Zhao, 2016). Mixed research methods combining teacher evaluation, self-assessment, peer review, writing portfolio, etc. are usually applied to classroom assessment. Therefore, the experimental results can hardly show which method is effective or not. As a result, the studies on different types of formative assessment have aroused more interest.

As a supplement to teacher evaluation, peer assessment adds new vitality to autonomous learning. Most of the research on peer assessment focused on writing ability, oral teaching, students' attitudes, etc., while only a few research consider different groupings in peer assessment and provide an explanation about how they group learners and why (Cai, 2011). For example, there are studies dividing students into high-level and low-level groups to examine the effectiveness of peer-to-peer evaluation but did not explain why they group learners in this way (Shen, 2013). Some studies have shown that students' different levels have an impact on the effectiveness of self-assessment or peer assessment. It is also found that peer assessment has a positive effect only on learners with high language proficiency but not on those with low proficiency (Liu, 2002; Sun & Li, 2015). That is to say, students with low proficiency level possibly cannot provide an effective assessment to their peer. Therefore, there might be a threshold of language proficiency for peer assessment to be effective which is suggested to be B1 and above (Cai, 2011).

In summary, there are studies showing that peer assessment is effective for students with a certain proficiency level but the validity of peer assessment varies with peers' proficiency. Students with high proficiency level may benefit a lot from peer assessment but those less proficient students do not. So it leads to a question, if different grouping methods will affect the validity of peer assessment, and if, in one peer group, students with different proficiency levels will benefit from peer assessment to a different degree. In order to maximize the validity of peer assessment, there is a need to investigate which grouping method is more effective. Therefore, this study will investigate if different ways of peer groupings have different effects on peer assessment. If yes, which grouping method is more effective.

## **2. Method**

This study aims at investigating the effect that different methods of peer grouping have on peer assessment. To this end, the study will examine the validity of three methods of peer grouping (i.e. high-low, high-high, low-low). Meanwhile, this study will try to justify the necessity of proper peer groupings in learning and teaching.

### *2.1 Research Questions*

Peer assessment in the classroom usually takes the form of group discussion. In natural classes, students' language proficiency varies. Therefore, the random grouping will possibly have the following combinations: groups with only learners of high proficiency levels (hereafter, HH group), groups with only learners of low proficiency level (hereafter, LL groups) and groups with both high and low levels (HL groups). This study tries to investigate the validity of different ways of grouping in peer assessment

To be specific, this study aims to answer the following questions:

Firstly, how do different ways of grouping influence writing improvement of the college students?

Secondly, what are the differences in feedback quality among different peer groups in terms of the adoption rate and accuracy rate?

Thirdly, how do the Chinese college learners perceive different ways of grouping in peer assessment?

### *2.2 Participants*

The participants in this study are 24 college students majoring in English including 7 undergraduates (with lower language proficiency level) and 17 first-year postgraduates (with higher language proficiency level). They have learned English for 12 years on average, with 10.5 years among undergraduate students and 13.5 years among postgraduate students. They are all willing and voluntary to participate actively in this study. They were assigned

to four kinds of groups, including LL, HL, HH groups. That is to say, two undergraduates randomly constituted an LL group, and two postgraduates formed a HH group. A postgraduate student and an undergraduate formed a HL group.

### 2.3 Research Design

This study collects both quantitative and qualitative data to investigate the validity of peer assessment.

Participants first received training on how to conduct peer assessment and practice it. First, the researchers gave a lecture on how to use the checklist to give feedback. The checklist consists of rating criteria on three aspects: language, structure and content. The training provided a well-written essay for every participant and a feedback example sheet for them to read. After that, participants were encouraged to rate the given essays and provide feedback. To produce effective feedback, the participants were asked to not only give their views on the three aspects of essays but also their reasons and suggestions on revision.

After that, all participants wrote an essay on the given topic *Majoring in humanities or science?* and handed it in anonymously. The researchers assigned the essays to the paired anonymous raters. The purpose of anonymity is to eliminate raters' subjective feelings towards their friends or more proficient participants. Then the participants evaluated the essay based on the checklist (rating criteria focusing on language, structure and content) prepared by the researcher, and gave written feedback.

After the assessment, the essays were given back to the authors. The authors checked each piece of feedback and decided whether they *agree with*, *disagree with* or *do not understand* the feedback. The researchers will calculate the accuracy/correctness rate and the adoption rate of the feedback.

Then the authors revised their writings according to the peer feedback and handed in both the original draft and the final version. The researchers will compare the scores of the two versions to see if there is improvement after peer assessment. Finally, all participants filled in a questionnaire about their perceptions on peer feedback and the validity of peer assessment.

The researchers transcribed all the writings into e-versions on the computer and rated them using online writing automatic grading system (<http://writing.bingoenglish.com/www/index.php/welcome>). This grading system was developed by experts from Zhejiang University and a company specializing in artificial intelligence. The rating is fast, and it can ensure the consistency and reliability of rating. This system is free online and it is widely used in China.

Then the researchers calculated the quantity, accuracy rate and adoption rate of the feedback in each essay as well as scores of the pre-revised and post-revised essays for data analysis.

## 3. Results

### 3.1 Scores of the Original Drafts and the Revised Versions

Descriptive statistical analysis was conducted to scores of the first and the revised draft. The results show that the average score of the revised draft is significantly higher than that of the first draft, which indicates an improvement in the quality of students' writings.

A paired sample t-test analysis (see Table 1) further evidences that the average score of the first and the revised draft is significantly different ( $df=23$ ,  $p=.002<.05$ ), which also reveals a significant improvement in the quality of students' writing between the two drafts.

Table 1 Scores of the two drafts and results of the paired sample t-test

Grouping	N	Average score of W1	Average score of W2	W2-W1	Sig(two-way)
LL	2	63.50	68.55	5.05	/
LH	5	69.94	70.34	0.4	
HL	5	70.60	72.04	1.44	
HH	12	74.00	77.01	3.01	
Total	24	71.57	73.88	2.31	

*Note.* HL and LH are the same group. But here, HL means students with higher proficiency (with a partner of lower proficiency); likewise, LH means students with lower proficiency (with a partner of higher proficiency).

In this sense, peer assessment is effective in terms of improving students' writing quality regardless of the

difference in the method of grouping. The difference of the average score shows that the writing quality improved most significantly in LL groups, followed by HH groups and relatively less significantly in the HL groups. One possible explanation is that students of the same language proficiency level may be able to provide feedback that is more understandable to their peers to facilitate learning. The increase of scores is small among lower proficient students in LH group. It is found that most feedback they received are correct but they refused to adopt those suggestions. It might be caused by the difficulty in comprehending the feedback beyond their competence. Apart from this, having a comparison between LH and HL groups, the improvement in scores of HL group increase more than LH group, which suggests that it seems easier for high-level students to understand and accept the feedback from low-level students while low-level students would lack the ability to judge the accuracy of the feedback from high-level students. Meanwhile, lacking relative knowledge in grammar, content and structure of low-level students, they were apt to provide high-level students with some simple and micro-level revising suggestions, mainly some grammatical errors. On the contrary, high-level students were inclined to give a comprehensive feedback (macro-level revision suggestions) on low-level students, including the unity, development, cohesion and coherence, structure, mechanism, wording.

### 3.2 Effectiveness of the Peer Feedback

As for the proportion of different types of feedback (focusing on language, structure or content), the low-level students pay more attention to language (68.18%) while the high-level students focus more on content and structure.

Accuracy rate and adoption rate of feedback are generally high among all the groups. In HL groups, accuracy rate and adoption rate are the lowest (83.75%, 58.58%), which reflects that the low-level students make more mistakes when they try to give feedback to the high-level students, while the high-level students can distinguish the feedback errors so as not to adopt the feedback. In high-level groups, the accuracy rate is high (95.40%), but the adoption rate is only 74.18%, which indicates that the high-level students were relatively more cautious when they receive feedback. In addition, Table 2 below shows that the feedback in LL group mainly concerns language (68.18%) and less deals with structure (18.12%) and content (13.64%). While HH groups pay comparatively more attention to content (26.63%) and structure (26.63%), meaning that compared with low-level students, high-level students would pay more attention to macro-level problems of writing and give high-level revising suggestions to peers.

In terms of the effectiveness of feedback, all the four grouping methods have presented a high accuracy rate and adoption rate, though the adoption rate of HL groups is comparatively low.

Table 2. Descriptive data of the peer feedback

	structure	content	language	accuracy rate	adoption rate
LL	18.18%	13.64%	68.18%	96.43%	83.93%
LH	19.15%	27.66%	53.19%	91.67%	82.70%
HL	18.18%	27.27%	54.55%	83.75%	58.58%
HH	26.63%	26.63%	46.63%	95.40%	74.18%

*Note.* accuracy rate=the number of accurate feedback/the number of all feedback; adoption rate=the number of feedback adopted/the number of all feedback.

### 3.3 Students' Perceptions on Peer Feedback

The questionnaire is composed of 5-point Likert-scale items on the effectiveness of feedback given to and received from peers. The response ranges from "strongly disagree" (1) to "strongly agree" (5). Reliability of the questionnaire is high (Cronbach alpha= .937). The average point of item responses is over 3.5, and the mean of total score is 3.79, which shows that students have positive views on peer assessment. Results of different groups show that the average score of low-level students' responses is higher than that of high-level students, reflecting that peer assessment is more desirable among low-level students.

Table 3. Results of the questionnaire survey

grouping	average	sd	Cronbach $\alpha$
LL	4.21	0.48	.937
LH	4.18	0.32	
HL	3.97	0.20	
HH	3.79	0.12	

#### 4. Discussion

To sum up, this study may draw the following conclusions: 1) Different combinations of grouping are all valid in peer assessment, but the validity varies. In this study, learners can improve their writings through peer feedback, especially for those groups of the same or similar proficiency level. This finding could be explained by the Zone of Proximal Development theory, which proposes that learners can acquire potential knowledge with assistance. But for those knowledge ways beyond learners' ability, it is possible that learners may feel hard to understand even with peer's assistance; 2) There is a high accuracy rate in students' feedback, but the accuracy rate and adoption rate of lower proficiency students in HL groups is relatively low, which means peer assessment in HL groups have a lower validity than in the other two grouping combination; 3) Students hold a positive view towards peer assessment and most of them believe peer feedback is helpful in promoting learning. It also shows that peer assessment is more popular among students with lower proficiency level.

The results of the study may shed some light on language teaching. In the classroom, teachers are encouraged to make the most of peer assessment to promote autonomous learning, especially in college. It could be seen that college students have the ability to provide effective assistance to their peers and they are willing to do peer assessment in the classroom. Peer assessment may allow students to take an active role in collaborative learning. Besides, when there are problems that cannot be tackled among peers, teachers could quickly recognize the main problems and provide effective guidance in learning. When conducting peer activity, one suggestion is that teachers could assign students with similar proficiency level into one group to achieve a better validity in peer assessment, for it will be much easier for students to acquire the knowledge that is within their potential developmental zone.

In all, this study tries to explore the effect of different groupings in peer assessment and it is found that different effect exists and that assigning students with similar proficiency level into one group can achieve a better validity in peer assessment. However, due to practical reasons, this study did not conduct in-depth interviews to investigate students' views on giving and receiving feedback. Future research might try to provide more empirical evidence to investigate peer assessment from learners' perspectives. Besides, this study investigates peer grouping in pairs. More research may be needed to examine the effectiveness of learning group of three and more.

#### Acknowledgments

This paper was based on the project "Validity of Different Peer Groupings in the Assessment of English Writings" (18GWXXM—48) funded by the Graduate Program of Scientific Research and Innovation in Guangdong University of Foreign Studies.

#### References

- Cai, J. (2011). A contrastive study of online peer feedback and online teacher feedback in teaching English writing to college students. *Foreign Language World*, 2, 65-72. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=37459710&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=37459710&from=Qikan_Search_Index)
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL journal*, 3(1), 5-19.
- Li, Q. (2010). The Impact of Portfolio-based Writing Assessment on EFL Writing Development of Chinese Learners. *Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press)*, 33(2). Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=1003766251&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=1003766251&from=Qikan_Search_Index)
- Li, Q. (2012). The status quo and future of formative assessment. *Foreign Language Testing and Teaching*, 3, 1-7, 26. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=42764882&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=42764882&from=Qikan_Search_Index)
- Li, Q. (2014). Constructing the Theoretical Framework of Formative Assessment. *Education Measurement and Evaluation (Theory)*, 4, 4-9.
- Li, Q., & Zeng, Y. (2008). Foreign language formative assessment: Validity and validation. *Foreign Language World*, 3, 82-90. Retrieved from

- [http://qikan.cqvip.com/Qikan/Article/Detail?id=27829405&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=27829405&from=Qikan_Search_Index)
- Li, X., & Ke, P. (2013). Efficacy of Peer Feedback in Translation Teaching. *Foreign Language Learning Theory and Practice*, 2, 83-88.
- Liu, J. (2002). Self-assessment of English Writing Skills by Chinese University Students. *Modern Foreign Language (Quarterly)*, 3, 241-249.
- Ma, J. (2012). *Peer feedback in college English writing*. Central China Normal University.
- Ma, Y. (2005). Research of Peer Cooperation Influenced by Theory of Zones of Proximal Development. *Shanghai English Studies*, 7, 39-41.
- Mendonca, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL quarterly*, 28(4), 745-769. <https://doi.org/10.2307/3587558>
- Paulus T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of second language writing*, 8(3), 265-289. [https://doi.org/10.1016/S1060-3743\(99\)80117-9](https://doi.org/10.1016/S1060-3743(99)80117-9)
- Shen, P. (2013). *A study of the effectiveness of combining peer assessment with teacher assessment on students' English writing proficiency in senior high schools in China*. Liaoning Normal University.
- Sun, X., & Li, Q. (2015). A comparative study of self, peer and teacher assessment in basic English writing class. *Chinese Foreign Language*, 1, 75-81. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=664965679&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=664965679&from=Qikan_Search_Index)
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of second language writing*, 9(2), 147-170. [https://doi.org/10.1016/S1060-3743\(00\)00022-9](https://doi.org/10.1016/S1060-3743(00)00022-9)
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press.
- Wang, H., & Zhen, F. (2008). Formative assessment based on language teaching and learning study and improvement. *Foreign Language Studies*, 3, 56-64, 112.
- Wen, Q. (2011). Formative assessment of the course on reading and evaluating research papers: Theory and practice. *Foreign Language Testing and Teaching*, 3, 33-38, 95. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=38627325&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=38627325&from=Qikan_Search_Index)
- Yang, H., & Wen, Q. (2013). A Review of Real-time Classroom Formative assessment: Reflections and Suggestions. *Foreign Language Teaching*, 3, 33-38. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=46885657&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=46885657&from=Qikan_Search_Index)
- Yang, H., & Wen, Q. (2014a). A Review of Real-time Classroom Formative assessment: Reflections and Suggestions. *Foreign Language Learning Theory and Practice*, 4, 41-45. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=46885657&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=46885657&from=Qikan_Search_Index)
- Yang, H., & Wen, Q. (2014a). Building “Moments of Contingency” in Real-Time Formative Assessment in Foreign Language Classrooms. *Foreign Language Education*, 4, 41-45. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=50139077&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=50139077&from=Qikan_Search_Index)
- Yang, H., & Wen, Q. (2014b). Building “Moments of Contingency” in Real-Time Formative Assessment in Foreign Language Classrooms. *Foreign Language Teaching and Research*, 3, 389-400, 480. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=50139077&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=50139077&from=Qikan_Search_Index)
- Yang, H., & Wen, Q. (2014b). Dynamic patterns of goals in real-time formative assessment in foreign language classrooms. *Foreign Language Teaching and Research*, 3, 389-400. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=49586484&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=49586484&from=Qikan_Search_Index)
- Zhao, W. (2016). A formative assessment in TCFL teachers' discourse: comparative study of novice and experienced teachers. *Foreign Language Teaching and Research*, 5, 1-10. Retrieved from [http://qikan.cqvip.com/Qikan/Article/Detail?id=669980916&from=Qikan\\_Search\\_Index](http://qikan.cqvip.com/Qikan/Article/Detail?id=669980916&from=Qikan_Search_Index)

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).