Article

Formative Assessment in Higher Education Classrooms: Second Language Writing Learning

Yujie Zhong
Guangdong University of Foreign Studies, China

Manzhen Yang*
Guangdong University of Foreign Studies South China Business College, China

Abstract

Although assessment for learning (AfL) has gradually become a dominant discourse in L2 writing classrooms in higher education throughout the world, in many cases, the quality of AfL implementation in real writing classrooms does not seem to be high enough to support students' authentic writing literacy development. Therefore, using 155 EFL first-year university students' written texts such as essays with written peer feedback, assessment sheets, reflective journals, and documents in e-portfolios as well as transcripts of interviews with four focal participants as the main data sources, this mixed-methods study seeks to investigate whether and how the implementation of AfL influences EFL students' learning in L2 argumentative writing. The quantitative results produced from analyzing the differences between students' writing scores before and after the intervention show that formative assessment significantly facilitates EFL students' learning of L2 argumentative writing skills. The qualitative accounts reveal a number of benefits that the students gain in L2 writing, such as enhancement of logical thinking ability, promotion of self-regulated learning, and improvement in peer feedback quality. The study also has implications associated with assessment design for L2 writing, teachers' roles in formative assessment, and teacher development programs.

Keywords

Assessment for learning, second language writing, blended learning, feedback, portfolios

1 Introduction

As indicated in Black and Wiliam's (1998) seminal research, assessment in education of all levels has gradually shifted its emphasis from students' final grades solely (i.e., summative assessment) to their ongoing progress made in the course of learning (i.e., formative assessment). Typically, in contemporary studies, formative assessment is primarily re-conceptualized as assessment for learning (AfL), which

*Corresponding Author
Address: Faculty of English Language and Culture, South China Business College, Guangdong University of Foreign Studies, Baiyun District, Guangzhou, 510545, China
Email: yangmzh@gdufs.edu.cn

refers to the use of "everyday practices" (Klenowski, 2009, p. 264) by teachers and students "to decide where the learners are in their learning, where they need to go and how best to get there" (Broadfoot et al., 2002, p. 2). In other words, when being exposed to a class that employs formative assessment approaches, students, teachers, and peers collaborate in establishing learning targets and criteria, then reflect upon and respond to various forms of feedback generated during the process with a concerted goal of enhancing ongoing teaching and learning (Stiggins, 2002). The advantages of this alternative to traditional large-scale testing are well-acknowledged. For example, it prompts students to be more active and self-regulated learners (Nicol & Macfarlane-Dick, 2006), encourages students' adoption of deep approaches to learning (Higher Education Academy, 2014), and promotes pedagogy, learning, and performance in high-stakes assessment in educational settings (Darling-Hammond & McCloskey, 2008).

In English as a foreign language (EFL) contexts, arduous efforts have been made to adopt and implement the principles and practices of AfL in second language (L2) writing instruction, especially in higher education (Aryadoust & Riazi, 2017). As key components encompassing AfL, the roles of feedback (including self, peer, and tutor feedback) and portfolios have been underscored in L2 writing, for they facilitate learner independence and reflections (Dann, 2014; Earl, 2013) as well as improve learners' L2 writing knowledge and skills (Diab, 2011). In addition, with technology evolving at an increasingly rapid rate, the combination of online and classroom instruction (i.e., blended learning) magnifies the advantages of AfL, guiding students to a more self-regulated pole of learning (Garrison & Vaughan, 2011). In this sense, the implementation of both AfL practices and blended learning can, to some extent, ameliorate the drawbacks of large-scale standardized tests that are product-based and teacher-dominated (e.g., timed impromptu essay testing; Black & Wiliam, 2018). However, despite the satisfactory achievements that AfL may bring ultimately, several syntheses argued that inaccurate and superficial understanding of AfL spelled its inadequate implementation at the classroom level (Aryadoust & Riazi, 2017; Black & Wiliam, 1998, 2018). Besides, a cohort of researchers have also manifested their doubts towards AfL and its effectiveness and provided some counter-evidence, which posed challenges to the validity and reliability in AfL (Bennett, 2011; Shute, 2008). Given the inconsistent conclusions that previous research has drawn, more empirical studies are needed in this research area.

This mixed-methods study attempts to investigate whether and how the adoption of formative assessment approaches in an EFL writing course at the university level will exert influence on students' learning in L2 argumentative writing. The assessment approaches included e-portfolios throughout the whole semester and self, peer, and teacher feedback on compositions. Moreover, apart from formal classroom instruction, students were also required to participate in online self-directed learning of a Massive Open Online Course (MOOC) named *Intermediate English Writing*. Drawing on the data collected through pre- and post-study writing scores, interviews, and reflection journals from 155 non-English major freshmen and four focal informants, this study seeks to contribute to the evidence of the usefulness of different AfL practices and to shed light on the issues about teaching design and methods, policy and teacher education in AfL contexts.

2 Literature Review

In this section, we will first discuss students' learning in assessment from a sociocultural perspective to lay the foundation of the present study. Then, related literature on AfL and students' learning of L2 writing in AfL-focused classrooms will be reviewed to justify the necessity of conducting further research and the methodological issues of the present study. Moreover, thanks to the affinities between feedback, portfolios, technology, and AfL, research in these areas will also be reviewed to explain the design of assessment approaches in the present study.

2.1 Sociocultural theory

Sociocultural theory (SCT) sees learning as a social process that goes beyond knowledge acquisition (Vygotsky, 1978). Under this paradigm, students' learning arising from activities is supposed to take place in real social contexts with interaction and/or collaboration with others (Driscoll, 2005). In other words, from a sociocultural perspective, learning occurs when individuals interact with others in social contexts as the first level, and when they internalize information into their mind as the second level. Therefore, learning is not a simple one-way process where the teacher imparts knowledge and the students passively receive it, but a social process in which students actively construct their knowledge through the use of sociocultural artifacts. Correspondingly, assessment as an integral part of any teaching and learning system should also entail rich interaction between teachers and students, students themselves, and even humans and culture (Benson, 2003). Ideally, the learner receiving it is expected to be an active participant who uses the information to improve their learning.

The instructor's role in AfL from a sociocultural point of view is generally seen as a facilitator, whose primary responsibility is to provide objective and formative feedback and scaffold students' learning to help them get their Zone of Proximal Development (ZPD; Lam, 2019), the gap between learners' actual development level and potential level characterized by external assistance (Vygotsky, 1978). While assessing, teachers or more competent peers act as scaffolds to help learners in need improve their learning, and they will gradually withdraw the aid as the learners become more capable and independent (Hyland & Hyland, 2006). Recently, Hanjani and Li (2014) confirmed that scaffolding does not necessarily comply with a top-down orientation; it also occurs between students of equal ability. Compared to summative assessment, AfL that aims at adapting teaching as well as enabling students to fully understand and enhance their learning (Klenowski, 2009) is more student-centered for it stresses not only the external requirements but also the learners' autonomy and initiative.

2.2 AfL and students' L2 writing in higher education

The past decades have witnessed a significant rise in research on AfL (Carless, 2017; Klenowski, 2009; Swaffield, 2011). As the pedagogical ideology transformed from a product-oriented to a more process-oriented one, the traditional summative assessment that aims to judge learning for accountability, ranking, or certification (Black et al., 2004) has gradually been replaced by its potent descendant, AfL, which is used as an equivalent of formative assessment in research in this field. It has gained increasing international prominence and an ever-growing number of education systems, as well as policymakers, have recognized the positive contributions it made to learners' development (Klenowski, 2009). As a result, AfL has been widely promulgated and employed in schools of all levels since the mid-1990s. Nevertheless, different teachers and researchers often assign different formulations to the term in their studies and practices, which might alter the original constructs that the term sought to promote (Gan et al., 2019). In this case, Black and Wiliam (1998) specified the concept of AfL in naturalistic classroom settings. The principles proposed by them serve as guidelines for avoiding misinterpretations. For example, according to Black and Wiliam (1998), practitioners of formative assessment should encourage productive classroom discussion, offer appropriate and timely teacher feedback, provide opportunities for students to learn from each other, and help enhance students' self-regulated learning awareness. Based on the above, in order to attain the full implementation of AfL, as appealed by some senior scholars (Black & Wiliam, 1998; Lam, 2016), more empirical investigations should be carried out to provide teachers with a precise and deep understanding of AfL as well as to exploit the potentials of AfL on boosting classroom-based teaching and learning.

Accordingly, AfL has become a dominant discourse in L2 writing classrooms in universities throughout the world (Birenbaum et al., 2015; Lam, 2016). Due to the experience of applications of AfL in first language (L1) writing classrooms where learners showed steady progress in their self-regulated

learning and cognitive and metacognitive capacity (Andrade & Boulay, 2003; Elbow & Belanoff, 1997; Yancey, 1998), AfL has gradually gained popularity in writing classes in English as a second language (ESL) or EFL contexts (Lam, 2016), with teachers and researchers in higher education being the most active and creative practitioners. Even though AfL is not considered a toolkit for raising students' grades (Zhang, 2020), it might have an impact on students' performance and learning in L2 writing. For instance, a significant number of studies have provided supporting evidence for the favoring effect of AfL on students' improvement of writing-related knowledge and skills by eliciting learners' self-accounts and illustrating medium to large effect sizes of increased achievements that the research participants had obtained (e.g., Brown, 2011; Love, 2009; Stiggins, 1999, 2006). In addition, scholars also proposed that students' autonomy and self-regulation have been fostered during the process (Black & Wiliam, 2018; Lee, 2011). However, in many cases, the quality of AfL implementation in real writing classrooms did not seem to be high enough to support students' authentic writing literacy development (Ranalli et al., 2018). As revealed in Mak and Lee's (2014) and Lam's (2019) research, teachers' rigid adoption of AfL in classes restricted its potential to bolster writing teaching and learning. Moreover, "AfL in the writing classroom was somewhat undermined against the backdrop of a test-driven culture" (Zhang, 2020, p. 2) which attaches great importance to language accuracy-based writing. Therefore, more research is warranted to expedite the regular conduct of formative assessment in L2 writing classes in higher education (Ranalli et al., 2018).

Methodologically, according to the literature review by Gikandi et al. (2011) that concerns theoretical and methodological issues of formative assessment in higher education, early research on this topic has generally adopted case study methodology because it was beneficial for the researchers to study the issue in situ with descriptive and interpretive data and in a more holistic way (Yin, 2003). However, as suggested by Krumsvik and Ludvigsen (2013), a mixed-methods research design may provide a more comprehensive picture of students' learning in a formative classroom because both processes and products of learning are contained. Moreover, Gikandi et al. (2011) have also mentioned that a mixed-methods approach is often required to establish the degree of validity and reliability in formative assessment. Hence, a mixed-methods research design should be applied to the study of AfL in L2 writing classrooms to attain richer results.

Therefore, in this study, we used a mixed-methods approach and adopted the definition of AfL by Black et al. (2004) that focuses on students' learning compared to teachers' teaching, for we regarded student performance in writing to be the center of our research:

> *Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence.* (p. 10)

Specifically, AfL in L2 writing is represented by the use of feedback (including self, peer, and teacher feedback) and student portfolios during a unit of instruction to enable students to be authentic writers equipped with sound writing and assessment literacy and decent self-regulated ability (Deeley, 2018; Lam, 2016; Zhang, 2020). Thus, the roles of feedback and portfolios are elaborated in the following section.

2.3 Feedback and portfolios in L2 writing

Previous literature has prioritized the status of feedback and portfolios as indispensable components of AfL (see Aryadoust & Riazi, 2017). On one hand, portfolios are typical symbols in formative classrooms. Writing portfolios stemmed from a call for a process approach to writing instruction, in which writers, writing processes, and higher-order cognitive competence become the emphases in class,

instead of higher grades to meet teachers' and parents' anticipation (Silva, 1990). Typically, in an average portfolio cycle in L2 writing classes, students are expected to use multiple sources for learning, compile the various learning records into different portfolios over time, and select the most characteristic work for showing the strengths and weaknesses of their writing development (Lam, 2014, 2016). Hamp-Lyons and Condon (2000) proposed a nine-step framework for portfolio assessment such as collection, reflection, selection, and delayed evaluation. By focusing more on learners' substantial gains in the process, portfolio assessment in writing classrooms has been proven to be an appropriate alternative to timed essay examinations (Lam, 2014; Nicol & Macfarlane-Dick, 2006).

On the other hand, AfL also necessitates effective feedback that makes a difference to the quality of writing, even to students' future learning. The definition of feedback as part of "a communication process through which learners enter into dialogues related to performance and standards" (Liu & Carless, 2006, p. 280) contains the notion that feedback involves several participants (peers and teachers) in the social construction of a text. However, considering the situations in higher education such as the mounting class sizes and numbers of lectures and research pressure on teachers, peer feedback is available more timely and in greater quantity in comparison with the more authoritative but delayed teacher suggestions (Nicol & Macfarlane-Dick, 2006). Moreover, students may incorporate more peer comments than teacher input in their revisions thanks to intelligibility and comprehensiveness (Xu & Liu, 2010; Zhao, 2010).

Nevertheless, especially in EFL contexts, people think teacher feedback is more useful than peer feedback. Evidence exists in Ruegg's (2018) experiment. For example, Ruegg (2018) divided numbers of Japanese university students into two groups (teacher feedback group vs. peer feedback group) to examine their changes in L2 writing self-efficacy over one academic year and found that students in the teacher feedback group outperformed the ones in the other group. Furthermore, self-feedback also plays a role in AfL in L2 writing classrooms, which facilitates the development of learner autonomy, promotes students' self-regulation in learning, and improves the writing quality (Diab, 2011; Suzuki, 2008). Therefore, given the respective benefits that self, peer, and teacher feedback possesses, a combination of the three sources of information may sprout the optimal achievements (Birjandi & Tamjid, 2012). Besides, Deeley (2018) suggested that drawing support from different types of technology nowadays could lay a promising future for effective AfL applications in higher education. Thus, incorporating blended learning into AfL might maximize their strengths to support and promote students' learning of writing over time.

2.4 Blended learning

Blended learning is defined as a learning environment where multiple sources of knowledge are delivered in various modes for the sake of facilitating students' learning (Garrison & Vaughan, 2011). Different delivery modes of specific information can play a catalytic role in exploiting the advantages of face-to-face classes and online instruction to the full. For instance, Adrianson and Hjelmquist (1999) found that classroom-based instruction was preferred when discussing and responding to ideas, whilst computer-mediated communication (CMC) was better for information-sharing. Moreover, research findings revealed that self-paced training offered through the use of technology has promoted students' learning (Park, 2011). Compared to regular classroom instruction at fixed times and locations, blended learning provides students with autonomy and flexibility to access the educational materials anytime anywhere, which might be more acceptable (Wang et al., 2009).

In this paper, blended learning is a combination of a twice-a-week face-to-face class with the courses and assignments delivered in the platform of ICourse, where Massive Open Online Courses (MOOCs) produced by expert teachers from well-known Chinese universities are available. In this case, students are expected to record what they have learned through the MOOC and compile them into e-portfolios. These practices help enhance their learning autonomy and are in turn embedded in the ecology of AfL in

the writing class (Deeley, 2018).

To sum up, even though AfL has been ratified to be a potent strategy to assist students' learning and further development, its implementation in higher education is unsatisfactory, which demands more empirical studies to be conducted to learn and improve the situation (Black & Wiliam, 1998; Lam, 2016; Ranalli et al., 2018). Because feedback and portfolios are vital constituents of AfL, formative assessment in L2 writing classes in higher education is framed by these substances (Aryadoust & Riazi, 2017). Besides, the collaboration of blended learning and AfL might offer new insights into the research field (Deeley, 2018). Therefore, this study that focuses on EFL university students' learning in an AfL and blended writing class seeks to investigate the following questions:

1. Whether the implementation of AfL can enhance students' learning in L2 writing?
2. How does the implementation of AfL influence students' L2 writing learning?

3 Research Design

Based on the methodological issues of formative assessment reviewed and the research questions listed above, a sequential explanatory mixed-methods design was applied in order to examine both quantitative and qualitative aspects of the impacts of AfL on students' learning of L2 writing skills (Tashakkori & Creswell, 2007). The use of quantitative numbers examined the effect of AfL on students' writing performance in a test-driven context. Meanwhile, the qualitative texts could better expound the quantitative data, thus helping the researchers to gain a more comprehensive understanding of students' learning in AfL-guided classrooms (Krumsvik & Ludvigsen, 2013).

3.1 Research context and participants

This mixed-methods research was conducted in an Integrated English class lasting for one semester, that is, 18 weeks, which was a compulsory course including reading, listening, speaking, and writing for first-year non-English major students at a university in southern China. Despite its integrated-skills nature, writing was a vital part of this class that the teacher emphasized. The course was taught twice a week, 80 minutes each, by a senior teacher (also the second author of this paper) who has more than 30 years of teaching experience and has devoted herself to the research of formative assessment, particularly peer feedback for many years. The site was chosen because formative assessment has been conducted in this kind of course in this school for years and the teacher herself is creative and adept at designing and implementing AfL in her instruction. In this case, the quality of AfL implementation could be ensured, which formed the premise of this research.

During the whole semester, the focus of the writing instruction has been zoomed in on one specific genre, argumentative writing. It forms a part of the College English Test (CET) Band-4 and Band-6 (a national English proficiency test for non-English major undergraduate students in China). Although the CET is neither mandatory nor related to students' obtaining of the baccalaureate degree, most college students in China wish to take the examination and strive for high marks to demonstrate their English proficiency because it might be beneficial to job-hunting upon graduation.

The first year at university is a critical moment for freshmen to understand the various requirements of the university. According to previous studies, formative assessment is an essential element that can assist students to achieve that aim (Poulos & Mahoney, 2008) and enhance their self-regulation (Black & Wiliam, 2018; Lee, 2011), both of which are important to students during their college life, even future development. From this aspect, formative assessment should have the most obvious influence on freshmen. Therefore, given the research questions raised, all the 155 first-year students (42 males and 113 females; mean age = 19.22 years) taught by Yang were recruited as the participants. They are native speakers of Chinese and most of them had learned English as a foreign language through

formal instruction for more than ten years. At the beginning of their undergraduate studies, they were assigned to three parallel intermediate-level classes based on their grades in a placement test. Among the 155 students, four focal informants were purposefully selected for further interviews based on their willingness to join the study, and ability to articulate their thoughts and their writing grades in the pre-test (Han & Hyland, 2015). Their demographic information is presented in Table 1.

Table 1

Background Information of the Focal Students

Name	Gender	Major	Year of college	Pre-test grades
Ling	Female	Accounting	First-year	12/15
Ting	Female	Accounting	First-year	10/15
Heng	Male	Computer Science	First-year	6/15
Qing	Female	Auditing	First-year	3/15

3.2 Procedure

In this course, the students were invited to set up study groups of two where they did peer review based on their preference and the grouping remained unchanged throughout the entire semester. Additionally, those students were in their second semester taking the writing class at the time of the current research. Hence, they were familiar with the formative assessment approaches employed during the process, which were operated on an assessment platform named F Platform for short (https://www.fifedu.com).

In the writing class, on one hand, each student was required to complete two argumentative writings with self, peer, and teacher feedback and the other two timed writings of the same genre that served as pre- and post-tests throughout the entire semester. The topics and rating scale of the timed writing referred to requirements of CET-4. When carrying out peer review, the students should work in pairs to comment on their partners' essays using Microsoft word processor based on the refined criteria in an assessment sheet, which was consistent with the one that was used in self and teacher editing in the semester. Then, after revisions, students' writings and written feedback were required to be submitted to the F Platform for teacher assessment.

On the other hand, the students were asked to attend online learning of a MOOC named *Intermediate English Writing*, participate in the speculative discussion of certain issues delivered in the class forum online, and collect and share the writing-related information they found in F Platform. At the end of the term, students should reflect on their learning on the word processor. These practices formed and enriched their e-portfolios.

During the 18 weeks' study, all the students were taught by the same English teacher with the same teaching materials, teaching procedures and were given the same writing assignments. The current study has received ethical approvals from the school, the teacher, and the students. The participants were assured of the confidentiality of the data collected, which would be exclusively used for academic research.

3.3 Data collection

3.3.1 Students' texts

The students' texts were collected as electronic documents from the F Platform with prior consent, including the two timed essays from all 155 students, along with the four informants' argumentative writings with peer feedback, assessment sheets, reflective journals, and documents in their e-portfolios.

The timed pre- and post-test essays serving as a measure of students' learning of L2 writing in an AfL and blended learning environment, the other peer-reviewed articles, assessment sheets, portfolio documents, and reflective journals were mainly used to unearth further information for the following interviews (Merriam, 1998).

The word count of the journals ranged from 194 to 713 Chinese characters, where students primarily recorded their thoughts, experiences, perceived strengths, and weaknesses of their learning during the semester. For example, they wrote about the knowledge and skills they had learned in writing classes, their feelings and thoughts about the classroom activities, and their experiences with peer comments. Besides, the e-portfolios contained their learning plans, learning records of the MOOC, their statement in the forum discussion, and the like.

3.3.2 Semi-structured interviews

The interview protocol was adapted from DeLuca et al.'s (2018) study and piloted with one student in the writing class. Then, the first author, Zhong, interviewed the target students in a one-on-one, face-to-face manner at the end of the research to understand and co-construct their experience and overall attitude towards the formative assessment and blended learning implemented during the term. The time of the interviews ranged from 30 to 45 minutes each. In the course of the semi-structured interviews, the researcher expected to elicit in-depth data on focal students' perceptions and practices of AfL, especially how they engaged in the process. Some pertinent points from the reflective journals and portfolios were also inquired and discussed together. In the study, all interviewees were interviewed in Chinese, a language they were most comfortable with and the interviews were audio-recorded with permission.

To enhance the trustworthiness of the findings (Lincoln & Guba, 1985), the recorded data were transcribed verbatim by Zhong and then sent to the respondents for verification. Extracts from both interview transcriptions and reflective journals were translated into English, which was checked again by Yang.

3.4 Data analysis

First, students' pre- and post-test essays were marked by Zhong and a Ph.D. candidate in education according to the standards of CET-4. The inter-rater reliability reached .928. All the disagreements were resolved through discussion and negotiation. The 155 student writing scores were entered into SPSS 24.0 to examine whether the students' writing performance improved under the influence of AfL. Descriptive data such as means and standard deviations (SD) and the paired t-test results were gleaned to see whether a significant difference existed in the marks of the two essays. The significance level was set at .05.

Second, to identify key constructs that influence students' L2 writing learning in an AfL context, transcripts of interviews and reflective journals were analyzed by Zhong and Yang jointly using thematic analysis (Braun & Clarke, 2006), and any disagreements in the coding were discussed and settled. The files were first read repeatedly, and some initial ideas were recorded. Special attention was paid to "anything pertinent to the research question or problem," but "new insights and observations that are not derived from the research question or literature review" were also noticed (Mackey & Gass, 2005, p. 241). After familiarizing with the transcripts, the researchers assigned initial codes to these extracts, which resulted in a long list of 127 codes ranging from "complicated writing tasks" and "MOOC learning" to "mastery of argumentative writing skills." Then, these codes were revised, removed, and grouped into three overarching themes over the recursive process of shuttling between empirical data and the existing literature in view of codes' internal homogeneity and external heterogeneity (Patton, 2015). For instance, codes such as "assess writing from multiple dimensions," "focus more on content and structures," "sense of collaboration," "reflections on my own writing," "conscious of reviewers and

receivers," and "mutual progress" were classified into the theme named "improvement in peer feedback quality."

4 Research Findings

4.1 Students' changes in L2 writing scores within an AfL context

As mentioned above, the writings of 155 students gleaned from three parallel classes were marked by two raters based on the assessment criteria of CET-4 with a maximum score of 15. The results of Pearson correlation analysis in the ratings of the essays indicated high reliability between the two raters (r=.928, p=.000). The analysis of students' writing grades in the pre-test showed that the mean scores for the three classes were 8.62 (SD=2.203), 9.18 (SD=2.297), and 8.50 (SD=2.964) respectively. The results of a one-way ANOVA showed no statistically significant difference in initial writing competence among the three classes (F=1.125, p=.327). Students' scores in the pre- and post-test writing tasks were also analyzed to examine whether a significant difference existed in the marks between the two tasks. The results of the paired samples *t*-test are presented in Table 2.

Table 2

Students' Writing Scores in Pre- and Post-test Writing Tasks

	N	M	SD	t	df
Pre-test	155	8.78	2.513	-6.412***	154
Post-test	155	10.16	1.895		

***p<.001

The table exhibits that the mean score of students' writing in the pre-test was 8.78 (SD=2.513), but it climbed to 10.16 (SD=1.895) in the post-test. To determine whether the difference in the mean scores of the argumentative writing before and after the intervention was statistically significant, a paired samples *t*-test was conducted. The results showed that students' writing scores in the two tests differed significantly (t=-6.412, df=154, p=.000). That is, students' average score of argumentative writing in the post-test was significantly higher than that in the pre-test. Therefore, the results suggest that the implementation of AfL and blended learning is conducive to enhancing students' L2 writing performance in light of their writing scores.

4.2 The influence of AfL on L2 writing learning

The four focal participants' self-accounts in the semi-structured interviews and their reflective journals were scrutinized to identify in what ways the formative assessment might affect the students' learning of L2 writing. Three overarching themes emerged from the analysis of the data.

4.2.1 Enhancement of logical thinking ability

Owing to the nature of the argumentative writing where the writers should persuade others to reach a consensus on diverse issues in a structured composition framework (a thesis statement, several arguments, and a conclusion), students need to develop their logical thinking ability which enables them to express their opinions explicitly and provide strong evidence to support themselves. According to the data collected in semi-structured interviews, the speculative discussion on the online class forum was conducive to developing students' logical thinking. For example, Qing stated:

> *The regular debates delivered in the class forum offered opportunities for us to organize our thoughts on specific topics. When I replied to the issues, I felt as if I was writing short argumentative writing and I needed to express myself clearly. I found that I could think about a few arguments to support my thesis statement in a more logical and systematic way in the last several discussions. Joining the online debates certainly played a role during the process.* (Interview with Qing)

When referring to the particular online discussion of assessing two sample essays in which students were invited to rate two pieces of argumentative writing and give reasons, Heng wrote in his reflection that "I was inspired by the model essays, as they prompted me to contemplate how to present my standpoints and ideas in a more organized and logical way" (Reflective journal of Heng). Besides, Ting also revealed her affirmative attitude towards the online discussion activity in her reflection and semi-structured interview.

> *This semester, the teacher pushed posts associated with the contents in our textbook on the F Platform. Through commenting on the issues, we could improve our English writing ability and exercise our thinking ability at the same time.* (Reflective journal of Ting)

> *The issues discussed in the class forum were all related to our attitudes towards something, which urged me to structure my scattered thoughts together because I wanted to convince those who disagreed with me. To this end, I should rearrange all kinds of reasons and evidence so that I could come to a logical conclusion that would also be acknowledged by others.* (Interview with Ting)

Moreover, the learning logs that the students recorded before they started to write the compositions were reported to make a difference in students' construction of their arguments.

> *Despite the fiddly steps we needed to complete when performing the writing tasks, I learned a lot. There are some useful guidelines on how to write an argumentative essay in the learning log. Responding to the questions in the learning log helped me realize my strengths and weaknesses in the course of writing and sort out my mind based on clear logic.* (Reflective journal of Ling)

In sum, analytical thinking ability is essential in argumentation due to the persuasive nature of this genre. Various formative assessment approaches adopted in writing classes such as online debates and learning logs assisted the participants to sharpen their logic in the process of L2 argumentative writing learning in the whole semester. Meanwhile, the enhancement of logical thinking ability, in turn, boosted students' development of argumentative writing skills.

4.2.2 Promotion of self-regulated learning

As mentioned earlier, apart from receiving the classroom instruction, students were required to attend online self-directed learning of a MOOC named *Intermediate English Writing* simultaneously. Students' statements in interviews and reflective journals suggested that this type of blended learning increased their participation in writing as well as facilitated the improvement of their writing competence.

> *The MOOC learning helped me obtain a deeper understanding of the methods and techniques for writing a good English composition...However, if we had merely taken the course without doing any practice in writing, it would have been hard for us to truly master the knowledge and skills. So, sometimes I would write a draft and then modify it based on the notes I had taken during MOOC learning.* (Interview with Ling)

This semester, we have learned the methods of writing an argumentative essay through MOOC, that is, Intermediate English Writing. The pre-recorded video course introduced tips for good writing in detail. Coupling with the argumentative writing task assigned by the teacher in class, I have made certain progress in my writing skills this semester. (Reflective journal of Ting)

In addition, Qing and Heng also commented on other AfL practices, namely, writing resources sharing and online speculative discussion conducted in writing teaching and learning all semester long. They reported that they were actively involved in the self-regulation of their work.

On one hand, I would make copies of some articles, videos, or audio files that might contribute to writing learning and upload them to the platform for sharing. On the other hand, I would also regularly review information shared by my classmates, especially the high-score model compositions of CET-4. Establishing a place for resource sharing was convenient and reciprocal for us to obtain extra lessons on L2 writing. (Interview with Qing)

I regarded it [speculative discussion] as a rehearsal of a real argumentative writing task. I usually forced myself to finish the reply within the allotted time in case I won't have enough time to complete my writing in CET-4. The classmates' replies also gave me new perspectives to demonstrate things. Therefore, I actively took part in online debates every time. (Interview with Heng)

In summary, the participants unveiled their affirmation of the significance of various formative assessment approaches employed in the writing class and their willingness to complete the writing tasks. The combination of online and classroom instruction accompanied by e-portfolios rendered the students considerable autonomy, fostered them to build up a sense of responsibility for their study, and directed them to a more self-regulated pole of writing learning.

4.2.3 Improvement in peer feedback quality

Another salient theme of the ways that formative assessment affects students' learning of L2 writing is the improvement in peer feedback quality. Peer feedback was valued in the writing instruction and constituted an indispensable part of AfL in L2 writing. Students in this class gave comments on their partners' essays using assessment sheets and received others' feedback on their writing in the meantime. The assessment sheet was considered a potent instrument that expedited the effectiveness of peer feedback on subsequent revisions.

The use of assessment sheets made a vast difference in helping us generate and receive more useful feedback on writing. We were asked to write down what parts of the essay we would like the reviewers to comment on at the top of the assessment sheet such as tense, vocabulary, content, or organization. By doing this, we usually focused more on the content and structure of the articles while reviewing, rather than looking for linguistic mistakes and giving superficial suggestions on revisions solely. (Interview with Heng)

Additionally, Ting also expressed her views in the semi-structured interviews. "When we held different opinions on certain problems, we would refer back to the criteria on the assessment sheet. It guided us to make better clarification and negotiation in the process of peer feedback" (Interview with Ting). In this sense, the assessment sheet instantiated the assessment criteria and helped the feedback givers and receivers reach a consensus on the standard.

Furthermore, Ling mentioned her gains from the peer editing tasks and highlighted the necessity of extending the dimensions of the comments in her reflective journal.

In the process of writing, we needed to evaluate and modify our own compositions as well

as those of our classmates. It enabled me to grasp the methods of evaluating compositions gradually and to realize the significance of revising articles. The feedback we provide for peers is not supposed to focus only on one side but on many perspectives. For example, I would check if my partner expressed her points of view clearly, offered sufficient evidence to support her standpoints, and developed the arguments in a logical way when I gave peer feedback. The more comprehensive feedback we receive, the better for us to improve our writing quality. (Reflective journal of Ling)

On the whole, an AfL context conduced to students' improvement of the quality of peer feedback considering the fact that the students have paid more attention to global problems such as content and structure in writing, instead of sticking to language errors. The use of a suitable assessment sheet could promote mutual understanding of the assessment criteria, which facilitated students' revisions on their compositions based on peer feedback.

5 Discussion

This mixed-methods study has examined the effect of the implementation of AfL on EFL first-year university students' learning of L2 argumentative writing in light of their writing scores as well as explored the benefits that students perceived during writing learning. The results of a paired samples *t*-test showed that students received significantly higher scores in the post-test after studying in an AfL context than they did in the pre-test at the very beginning of the research. Additionally, through analyzing data collected from semi-structured interviews and reflective journals of the focal informants, the findings have demonstrated that AfL practices played an active role in enhancing students' logical thinking ability, promoting their self-regulated learning, and improving their peer feedback quality.

Echoing previous research (e.g., Brown, 2011; Love, 2009; Stiggins, 1999, 2006), the current study has further verified that AfL contributes to EFL students' learning of writing knowledge and skills from different perspectives as revealed by students' remarks in interviews and reflective journals. A statistically significant difference in students' writing scores before and after the research has also evinced the positive influence of formative assessment on EFL students' learning of L2 writing. In the writing classes, the students were given enough chances to interact with others in multifaceted contexts such as classroom instruction, peer feedback, and online forum debates. During these processes, the teacher and peers in the classroom all acted as scaffolds for each other to improve teaching and learning (Hanjani & Li, 2014). As the focal participants implied in the interview that the important roles that teacher guidance and peer assistance played in their learning, formative assessment is rich in interpersonal communication, making the assessment process a two-way system (Nicol & Macfarlane-Dick, 2006). In this case, all the people involved are required to be active participants and use the information to alter their ZPD, which contributes to students' improvement in logical thinking, self-regulation, and feedback quality. Besides, as SCT indicates that socio-cultural artifacts can also make a difference, disparate formative instruments such as learning logs, assessment sheets, and MOOCs constitute boosters of students' learning development. For example, the metacognitive and cognitive guiding questions in the learning log were reported to be hints for students' generation and arrangement of their ideas and to be the references when they were required to express their opinions on controversial issues, which helped students to enhance their logical thinking ability.

Furthermore, even though students' learning is prioritized in the present study, the findings have also shed light on the teacher's teaching under an AfL paradigm whose crux is the student-centered principle, which implies "teachers' adaptation of the spirit of AfL rather than solely the procedural practices" (DeLuca et al., 2012, p. 24). It is easy for teachers to be influenced by the test-driven culture (Zhang, 2020), neglect students' needs (Lam, 2019), and fall into the trap of "simplifying" AfL (Gan et

al., 2019), taking for granted that formative assessment is to offer students the information about their learning, not just a score. However, our responsibility as teachers does not end unless the students truly understand the purpose and means of formative assessment and then take action (Carless, 2017). Thus, teachers are supposed to be facilitators of students' learning in formative assessment. In the current research, as indicated in the students' remarks, although some students might feel tedious when they first engaged in various activities such as speculative discussion, resource sharing, and self and peer feedback, they changed their minds as they gradually found the benefits and necessity of conducting this kind of assessment. For instance, when they realized that they could organize the essay more logically and that they could provide more constructive and all-around feedback through social interaction, they derived a sense of fulfillment. Meanwhile, their initiative was aroused for further action. Student-centered assessment can encourage meaningful dialogue, increase collaboration, and form a community of shared responsibility (Klenowski, 2009).

The present study attempts to demonstrate an alternative way to conduct the student-centered assessment in L2 writing classrooms. Within a writing cycle in this research, diverse formative components functioned before, during, and after writing. For instance, before writing the compositions, students were required to fill in a learning log where some cognitive and metacognitive questions that guided students to structure their essays were presented. As indicated in students' accounts, the learning log, a part of students' portfolios, helped sharpen their logic (Saadé et al., 2012), which was a positive catalyst to the development of students' cognitive and metacognitive capacity (Andrade & Boulay, 2003; Elbow & Belanoff, 1997; Yancey, 1998) and writing literacy (Ranalli et al., 2018).

In addition, in completing a writing task, the students provided and received feedback for and from different people, such as the teacher, the peers, and the students themselves. However, peer feedback generally took up a large proportion of the resources for the following revisions due to its accessibility (Nicol & Macfarlane-Dick, 2006) and intelligibility (Xu & Liu, 2010; Zhao, 2010). Through giving and receiving feedback on writing, students' participation in writing was improved (Liu & Carless, 2006), which in turn made them take an active role in the process (Zhu & Carless, 2018). The qualitative materials indicate that immersing in an AfL context with varied formative instruments such as learning logs, assessment sheets, and MOOCs can allow students to focus on global areas of writing (Lam, 2014; Mak & Lee, 2014) and enhance their peer feedback practices (Diab, 2011). Besides, students may adopt some cognitive or metacognitive strategies when they generate feedback for their peers (Moretti et al., 2015). For example, they were more likely to reflect on their writing and discern the competence gaps when reviewing the compositions written by peers, which helped to alter their ZPD, enhance self-regulation awareness (Lam, 2014), and promote cognitive and metacognitive development (Andrade & Boulay, 2003; Elbow & Belanoff, 1997; Yancey, 1998).

Furthermore, some formative practices, that is, resource sharing, speculative discussion, and MOOC learning were interspersed throughout the whole semester. Corresponding to Deeley's (2018) assertion that the use of CMC can be "beneficial in facilitating effective assessment for learning and feedback in higher education," (p. 439) this study has also instantiated that a combination of online writing instruction (i.e., MOOC learning) and classroom teaching promoted self-regulation (Garrison & Vaughan, 2011) and writing skill development. Moreover, the online forums for resource sharing and analytical debates offered opportunities for students to express their views freely over some controversial topics, exchange ideas with each other, and subtly establish a reciprocal learning community for mutual progress (Carless, 2013). Accordingly, students' logical thinking ability (Saadé et al., 2012) and a sense of responsibility for their own learning (Lam, 2014) were fostered during the process simultaneously. In this case, portfolios that showcased the strengths and weaknesses of learners' writing development over a term were shaped, which was consistent with the portfolio assessment framework put forward by Hamp-Lyons and Condon (2000).

6 Conclusion and Implications

The present study reveals that the implementation of formative assessment in higher education positively mediates EFL students' learning of L2 argumentative writing. Incorporating diverse formative assessment tools such as feedback on compositions and portfolios, which includes MOOC learning records, learning logs, reflective journals, and discussion and resource sharing records, students' writing knowledge and skills are improved in an AfL oriented context. Further, the study shows that AfL is beneficial in enhancing students' logical thinking ability, promoting their self-regulation, and improving their feedback quality, all of which in turn contribute to L2 writing learning.

Before discussing the implications of the study, it is important to mention its limitations. One limitation involves the generality of the findings. In the present study, we only analyzed the self-revealing account of the four focal participants while investigating the specific ways that AfL influenced L2 writing. Consequently, it remains unanswered whether the factors would be the same when a large cohort of participants with various learning backgrounds are inquired. Another limitation is that we have only appraised students' writing learning in terms of their progress in writing scores. There is a need to look into students' changes from other perspectives, such as behavior and emotion.

Nonetheless, important implications can be drawn from the study. First, teachers of L2 writing are recommended to adopt various formative tasks and assessment approaches that facilitate students' ongoing progress in learning (Ranalli et al., 2018). For example, establishing class forums for exchanging ideas and initiating discussions, or conducting self and peer feedback on writing with auxiliary instruments such as learning logs and assessment sheets can be feasible means to promote students' learning development. Moreover, to ensure the effectiveness of formative assessment, teachers should readjust their roles in classrooms and devolve autonomy and initiative to students themselves to construct a student-centered learning environment jointly (Zhu & Carless, 2018). At a time when the requirement for foreign language ability has been rising whilst college English classes are getting squeezed, the traditional college English curriculum design and the cramming method of teaching are inappropriate in cultivating self-regulated learners in formative contexts. Furthermore, institutional support (Lam, 2019; Mak & Lee, 2014) is supposed to go hand in hand with the promulgation of AfL in higher education. In order to improve the quality of AfL implementation in real writing classrooms, teacher development programs, especially those related to evaluation and testing should be widely carried out and diffused. In terms of research, more qualitative studies can be conducted to explore students' L2 writing learning in AfL contexts that are assessed from multiple dimensions. Besides, longitudinal investigations can also be launched to unearth students' learning trajectories with formative assessment over a period.

References

Adrianson, L., & Hjelmquist, E. (1999). Group processes in solving two problems: Face-to-face and computer-mediated communication. *Behaviour & Information Technology, 18*(3), 179-198. https://doi.org/10.1080/014492999119075

Andrade, H., & Boulay, B. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research, 97*(1), 21-30. https://doi.org/10.1080/00220670309596625

Aryadoust, V., & Riazi, M. (2017). Future directions for assessing for learning in second language writing research: Epilogue to the special issue. *Educational Psychology, 37*(1), 82-89. https://doi.org/10.1080/01443410.2016.1248134

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Benson, A. D. (2003). Assessing participant learning in online environments. *New Directions for Adult*

and Continuing Education, 100, 69-78. https://doi.org/10.1002/ace.120

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., … Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education, 13*(1), 117-140. https://doi.org/10.1177/1478210314566733

Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education, 37*(5), 513-533. https://doi.org/10.1080/02602938.2010.549204

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8-21. https://doi.org/10.1177/003172170408600105

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge University Press.

Brown, G. (2011). Self-regulation of assessment beliefs and attitudes: A review of the students' conceptions of assessment inventory. *Educational Psychology, 31*(6), 731-748. https://doi.org/10.1080/01443410.2011.599836

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In D. Bond & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 90-103). Routledge.

Carless, D. (2017). Scaling up assessment for learning: Progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan., & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 3-17). Springer.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149-166. https://doi.org/10.1080/0969594X.2014.898128

Darling-Hammond, L., & McCloskey, L. (2008). Assessment for learning around the world: What would it mean to be internationally competitive? *Phi Delta Kappan, 90*(4), 263-272. https://doi.org/10.1177/003172170809000407

Deeley, S. J. (2018). Using technology to facilitate effective assessment for learning and feedback in higher education. *Assessment & Evaluation in Higher Education, 43*(3), 439-448. https://doi.org/10.1080/02602938.2017.1356906

DeLuca, C, Chapman-Chin, A. E. A., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *The Curriculum Journal, 29*(1), 77-94. https://doi.org/10.1080/09585176.2017.1401550

DeLuca, C., Luu, K., Sun, Y., & Klinger, D. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters, 4*, 5-29. https://doi.org/10.18296/am.0104

Diab, N. M. (2011). Assessing the relationship between different types of student feedback and the quality of revised writing. *Assessing Writing, 16*(4), 274-292. https://doi.org/10.1016/j.asw.2011.08.001

Driscoll, M. P. (2005). *Psychology of learning for instruction* (3rd edition). Allyn and Bacon.

Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd edition). Corwin.

Elbow, P., & Belanoff, P. (1997). Reflections on an explosion: Portfolios in the 90s and beyond. In K. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 21-33). Utah State University Press. https://doi.org/10.2307/j.ctt46nxw3.4

Gan, Z., He, J., & Mu, K. (2019). Development and validation of the assessment for learning experience inventory (AFLEI) in Chinese higher education. *Asia-Pacific Education Research, 28*(5), 371-385. https://doi.org/10.1007/s40299-019-00435-7

Garrison, D. R., & Vaughan, N. D. (2011). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass.

Gikandi, J., Morrow, D., & Davis, N. (2011). Online formative assessment in higher education: A review of the literature. *Computers and Education, 57*(4), 2333-2351. https://doi.org/10.1016/j.compedu.2011.06.004

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio principles for practice, theory, and research* (Written language series). Hampton Press.

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*(4), 31-44. https://doi.org/10.1016/j.jslw.2015.08.002

Hanjani, A., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System, 44*, 101-114. https://doi.org/10.1016/j.system.2014.03.004

Higher Education Academy. (2014). *Framework for partnership in learning and teaching in higher education*. The Higher Education Academy.

Hyland, K., & Hyland, F. (Eds.). (2006). *Feedback in second language writing: Contexts and issues*. Cambridge University Press.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263-268. https://doi.org/10.1080/09695940903319646

Krumsvik, R., & Ludvigsen, K. (2013). Theoretical and methodological issues of formative e-assessment in plenary lectures. *International Journal of Pedagogies & Learning, 8*(2), 78-92. https://doi.org/10.5172/ijpl.2013.8.2.78

Lam, R. (2014). Promoting self-regulated learning through portfolio assessment: Testimony and recommendations. *Assessment & Evaluation in Higher Education, 39*(6), 699-714. https://doi.org/10.1080/02602938.2013.862211

Lam, R. (2016). Assessment as learning: Examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in Higher Education, 41*(11), 1900-1917. https://doi.org/10.1080/03075079.2014.999317

Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System, 81*, 78-89. https://doi.org/10.1016/j.system.2019.01.006

Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching, 5*(1), 19-33. https://doi.org/10.1080/17501229.2010.502232

Lincoln, S., & Guba, G. (1985). *Naturalistic Inquiry*. SAGE.

Liu, N., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290. https://doi.org/10.1080/13562510600680582

Love, N. (2009). Building a high-performance data culture. In N. Love (Ed.), *Using data to improve learning for all: A collaborative inquiry approach* (pp. 2-24). Corwin.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum.

Mak, P., & Lee, I. (2014). Implementing assessment for learning in L2 writing: An activity theory perspective. *System, 47*, 73-87. https://doi.org/10.1016/j.system.2014.09.018

Merriam, S. B. (1998). *Qualitative research and case study applications in education: Revised and expanded from "case study research in education."* Jossey-Bass.

Moretti, G., Giuliani, A., & Morini, A. (2015). Flexible and dialogic instructional strategies and formative feedback: An observational practice in Italian high schools. *Proceedings of International Conference of Education, Research and Innovation 2015*, 8229-8236. https://www.researchgate.net/publication/316191933_FLEXIBLE_AND_DIALOGIC_INSTRUCTIONAL_STRATEGIES_AND_FORMATIVE_FEEDBACK_AN_OBSERVATIONAL_RESEARCH_ON_THE_EFFICACY_OF_ASSESSMENT_PRACTICES_IN_ITALIAN_HIGH_SCHOOLS

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Park, Y. (2011). A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types. *International Review of Research in Open and Distance Learning, 12*(2), 78-102. http://www.irrodl.org/index.php/irrodl/article/view/791/1699

Patton, M. (2015). *Qualitative research and evaluation methods: Integrating theory and practice*. SAGE.

Poulos, A., & Mahoney, M. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143-154. https://doi.org/10.1080/02602930601127869

Ranalli, J., Feng, H., & Chukharev-Hudilainen, E. (2018). Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing, 36*, 77-89. https://doi.org/10.1016/j.asw.2018.03.007

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *Language Learning Journal, 46*(2), 87-102. https://doi.org/10.1080/09571736.2014.958190

Saadé, R., Morin, D., & Thomas, J. (2012). Critical thinking in E-learning environments. *Computers in Human Behavior, 28*(5), 1608-1617. https://doi.org/10.1016/j.chb.2012.03.025

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11-23). Cambridge University Press.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27. https://doi.org/10.1111/j.1745-3992.1999.tb00004.x

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-765. https://www.jstor.org/stable/20440249

Stiggins, R. J. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Educational Testing Service.

Suzuki, M. (2008). Japanese learners' self revisions and peer revisions of their written compositions in English. *TESOL Quarterly, 42*(2), 209–233. https://doi.org/10.1002/j.1545-7249.2008.tb00116.x

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice, 18*(4), 433-449. https://doi.org/10.1080/0969594X.2011.582838

Tashakkori, A., & Creswell, J. W. (2007). Editorial: Exploring the nature of research questions in mixed methods research. *Journal of Mixed Methods Research, 1*(3), 207-211. https://doi.org/10.1177/1558689807302814

Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wang, M., Shen, R., Novak, D., & Pan, X. (2009). The impact of mobile learning on students' learning behaviours and performance: Report from a large blended classroom. *British Journal of Educational Technology, 40*(4), 673-695. https://doi.org/10.1111/j.1467-8535.2008.00846.x

Xu, Y. & Liu, J. (2010). An investigation into anonymous peer feedback. *Foreign Language Teaching and Practice, 3*, 44-49.

Yancey, K. B. (1998). *Reflection in the writing classroom*. Utah State University Press.

Zhang, X. (2020). Assessment for learning in constrained contexts: How does the teacher's self-directed development play out? *Studies in Educational Evaluation, 66*, 1-9. https://doi.org/10.1016/j.stue-duc.2020.100909

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing, 15*(1), 3-17. https://doi.org/10.1016/j.asw.2010.01.002

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development, 37*(4), 883-897. https://doi.org/10.1080/07294360.2018.1446417

Yujie Zhong is a postgraduate student at Guangdong University of Foreign Studies. Her research interests include formative assessment and peer feedback on L2 writing.

Manzhen Yang is a professor at Guangdong University of Foreign Studies South China Business College in Guangzhou, China. Her research focuses on language testing and assessment, formative assessment, and teacher development.