

The Effectiveness of Dynamic Assessment in Linguistic Accuracy in EFL Writing: An Investigation Assisted by Online Scoring Systems

Liuyan XIAN*

Lingnan Normal University, Hong Kong Polytechnic University

Received 01 April 2020 Accepted 13 July 2020

Abstract

From the perspective of Vygotsky's Sociocultural Theory and Zone of Proximal Development (ZPD), dynamic assessment (DA) highlights the role of mediation in the learning process by unveiling the distance between actual learning proficiency and potential learning development. DA is more helpful in fostering students' ability to solve problems independently and in facilitating future tutoring. However, due to its large consumption in time, a vast majority of studies on DA target at a limited scale. To remove this limitation from the implementation of DA, this paper attempts to apply interventionist DA with the support of online scoring systems in EFL writing. Targeting at 44 Chinese L2 English learners, this experimental research mainly adopts quantitative methods with qualitative methods as a supplement to explore their improvement in linguistic accuracy in English writing. It is found that both non-dynamic assessment (NDA) and DA helped participants to improve their accuracy, but only in the DA group observed a remarkable statistical significance. That is, DA is more effective in promoting learning outcomes. Besides, ZPD is predictive for future learning achievement.

Keywords: *Zone of Proximal Development, Dynamic Assessment, Linguistic Accuracy, Mediation*

Introduction

Dynamic assessment (hereafter: DA) is derived from Vygotsky's Sociocultural Theory (SCT). Catering to learners' different needs, it simultaneously evaluates learning outcomes and promotes learning development through mediation (Lantolf & Poehner, 2007). Distinguished from traditional static assessments (roughly refer to formative assessment and summative

assessment), DA not only assesses the actual learning performance but also predicts learners' learning performance in the future (Lantolf & Poehner, 2007). Mediation is a core concept in DA since it underlines modifiability and instructional suggestions to the improvisation of learning accomplishment (Lidz, 1991). To be more straightforward, DA incorporates assessment and mediation into the process of learning, and students are highly engaged as a vital component of learning. However, DA requires tailored prompts or interaction between the mediator and the learners, inducing labor intensiveness, and time demand (Antón, 2009; Lantolf & Poehner, 2007; Yang & Qian, 2019). Thus, it is a challenge to apply DA on a large scale in classroom teaching. To fulfill this gap, the paper, targeting at minimizing tutors' workload, explores DA with supports from the online scoring system. Moreover, albeit few studies addressing the effectiveness of computerized DA (e.g., Poehner & Lantolf, 2013; Poehner, Zhang & Lu, 2015; Yang & Qian, 2019; Zhang & Lu, 2019) or group DA (Afshari, Amirian, & Tavakoli, 2020; Shabani, 2018), most of them focus on listening or reading. Meanwhile, in writing, it is still a challenge to expand the scope of applying DA because of the changeability and subjectivity in the assessment criterion. Therefore, studies on web-based DA in writing are quite rare (e.g., Ahmadi & Besharati, 2017; Alderson & Huhta, 2005; Davoudi & Ataie-Tabar, 2015; Mehri Kamrood, Davoudi, Ghaniabadi, & Amirian, 2019; Vakili & Ebadi, 2019), so more empirical research is needed to scaffold a wide application of web-supported DA in writing teaching and learning. The paper is designed to explore, with online scoring systems' support, DA's efficacy in English writing and its capacity in predicting future learning performance in an EFL context.

Background

Vygotsky's Sociocultural Theory (SCT), especially the Zone of Proximal Development (ZPD), underpins DA. SCT is constructed on the basis that social and cultural interactions frame human cognition (Watson-Gegeo, 2004). Thus, cognition, meaning, and interaction are closely related to this theory (Lantolf, Thorne, & Poehner, 2015). Different from animals, human beings' interactions with the environment take place in a mediated way rather than a direct way. That is, new cognition occurs intermentally and intramentally (Vygotsky, 1978). In intermental plane, entities exchange information via symbols or languages through interactions, collaborations or mediation. The information is collected and internalized by the agent, and the intramental plane works. Thus, the agent can understand the information and apply it in a brand-new situation to solve problems independently. This cognition process is reflected in ZPD as "the distance between the actual developmental level as determined by independent problem solving and level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978: p. 86). In ZPD, collaboration, a fundamental and vital concept in cognition (Lantolf, 2000), indicates to mediation and interaction. An ideal way for learners forward from the actual developmental level to the potential developmental level is to provide mediation through interaction to filtered and framed learning experiences.

DA extends the conviction of learning by interaction proposed by SCT and ZPD to assessment. Traditionally, assessments evaluate learners statically by focusing on past performance at a time point. The assessors are expected to scale learners' performance from a neutral and objective standing and endow students with little or no feedback in terms of their achievements (Sternberg & Grigorenko, 2002). DA, however, distinguishes itself from traditional assessments in that it outstands future achievement and learning in the process of assessment, with tailored feedback and mediation by examiners' intervention (Sternberg & Grigorenko, 2002). To demonstrate, the distinction between static assessment and DA lies in whether mediation is embedded in the process of assessment (Lantolf & Poehner, 2007).

Interactionist and interventionist are two mainstreams in DA (Lantolf & Poehner, 2007). In terms of interactionist DA, mediation is embedded in the process of interaction between the tutors and the learners, so this approach is also named as *teaching in assessment* (Allal & Ducrey, 2000). Due to its attachment to the classroom context, interactionist DA shares some similarities with formative assessment (Lantolf & Poehner, 2007). Meanwhile, the mediation from an interventionist view, possesses a set of standards to pursue a fast pace of learning (Brown & Ferrara, 1985). Differing from interactionist DA, it is more likely to be applied on a large scale in education. Accordingly, it is comparable with summative assessments (Lantolf & Poehner, 2007) and is called *assessment in teaching* (Allal & Ducrey, 2000). Attempting to apply DA into a large-scale education context, this paper is underpinned by interventionist DA; hence, it draws a further discussion in the following paragraph.

Interventionist DA can be further classified into the sandwich format and cake format (Sternberg & Grigorenko, 2002). The former, with a traditional pretest-intervention-posttest pattern, calculates the gains between the initial test and the last test to evaluate the efficacy of DA on learning outcomes. That is, intervention is separated from the assessment. The latter format, however, integrates mediation into assessment by offering prompts implicitly to explicitly and generates an individual profile for each learner with a set of grades, specifying the mediation and performance information. The sandwich format coincides with quantitative studies, whereas the cake format corresponds to qualitative researches traditionally.

Literature Review

In language education, DA has been applied into different aspects: listening, speaking, reading and writing, but listening and reading have attracted a large number of studies (e.g., Ableeva & Lantolf, 2011; Guterman, 2002; Lantolf, 2009; Mehri Kamrood et al., 2019; Poehner & Lantolf, 2013; Poehner & Leontjev, 2020; Yang & Qian, 2019). In writing research, DA studies can be reviewed in terms of interventionist DA (e.g., Davoudi & Ataie-Tabar, 2015; Derakhshi, 2019; Hidri, 2019; Vakili & Ebadi, 2019; Xiaoxiao & Yan 2010; Zhang, 2013; Zarbafian, Abbasian, Mohseni & Baradaran, 2020), interactionist DA (e.g., Afshari et al., 2020; Antón, 2009; Ahmadi & Besharati, 2017), and a comparative study of different DA modes (e.g., Hassaskhah & Haghparast, 2012). However, due to the space limitation, and this paper is built on an interventionist DA, only relevant studies are discussed.

Under the framework of interventionist DA, Xiaoxiao and Yan (2010) expanded the scale from a limited number to an EFL classroom context to explore the implementation of DA in line with students' writing ability and learners' motivation in writing. This case study adopted a simultaneous one-to-many interactive pattern in mediation and reported an improvement in meaning, content and confidence in writing and high satisfaction from participants. However, this study targets a few samples, and the findings are hard to generalize in a different teaching environment. Zhang (2013) specified Xiaoxiao and Yan's (2010) writing process and introduced an online tool and peer review to facilitate DA's application to a larger scale. Zhang's (2013) design remarkably reduces teachers' workload and it is suitable to apply in a large-scale class. But Zhang's design is a proposal instead of empirical research; thus, data are needed to provide evidence to its efficacy. Davoudi and Ataie-Tabar's (2015) study, based on computerized dynamic assessment (CDA), provided empirical evidence to Zhang's (2013) proposal. Embedding CDA in the process of writing with an interactive and strategy-based learning context, this paper proposed that writings have been improved. Besides, those language learners with low proficiency gained more benefits than others in the experiment. Derakhshi (2019) reported similar results that interventionist DA played a positive role in test performance and linguistic accuracy and Zarbafian et al. (2020) pointed out that DA was a useful tool to increase learners' engagement in critical thinking. Vakili and Ebadi (2019), with a qualitative study, elaborated on the effects of face-to-face DA and computerized DA that the former induced collaboration in learning while the latter encouraged student engagement and led to self-regulation. Hidri (2019) articulated a different voice that the misuse of mediation strategies may fail to activate the prior knowledge and test-taking strategies to apply the knowledge in a new environment. Hassaskhah and Haghparast (2012) compared two modes of DA and stated that both interventionist and interactionist DA worked more effectively than traditional assessment.

Though studies have demonstrated that interventionist DA can facilitate writing in linguistic accuracy, overall quality, student engagement and collaboration, they fail to make use of the popular application of online scoring systems to reduce assessors' workload to expand DA to a larger scale.

The Study

Since the previous studies on DA in EFL writing are limited, and few studies make use of the support of online scoring systems to promote DA to large scope, the experimental study integrates the qualitative method into quantitative methods to explore the role of DA in EFL writing with the help of online scoring systems to seek answers to the following three questions:

1. In the linguistic accuracy of writing, what are respective performances of participants' in the non-dynamic assessment group (NDA group) and the dynamic assessment group (DA group) in terms of the pretest, immediate posttest and transfer posttest?
2. Which group plays a better role in improving linguistic accuracy in writing?
3. Can ZPD in pretest predict participants' future performance?

Based on the findings in Davoudi and Ataie-Tabar (2015), Hassaskhah and Haghparast (2012), Vakili and Ebadi (2019), Xiaoxiao and Yan (2010), and Zarbafian et al. (2020), it is hypothesized that in both NDA and DA group, a reduction of linguistic errors would be observed, but the DA group works more effectively. And based on Lantolf and Poehner (2007), ZPD can predict participants' future learning outcomes.

Research Design

Participants

The target participants were sophomore English-majors in the School of Foreign Studies of Lingnan Normal University. The researchers informed all the students in the second year in the school of the writing research, and they were all invited to attend this experiment. Students in other majors and in other grades were excluded because the first criterion to judge students' comparability was their gains in Comprehensive English III, a major course bearing 4 credits in the semester for English majors. In the final test of Comprehensive English III, students were assessed in line with vocabulary, grammar, reading and writing.

Forty-four students responded to the call, 5 of whom were male and 39 female, ranging from 18 years old to 22 ($M=20.25$). All the subjects spoke Cantonese or Chinese Mandarin as L1 and English as their first foreign language. In the meantime, all of them also learned another foreign language, such as French, Japanese, Korean, and others. Up to the experiment time, they had accepted formal English instruction for at least 9 years, and at most 15 years ($M=11.95$). Only four participants had traveled or studied in an English-speaking country for no more than three months. On average, they took English-related courses for 13.83 hours per week, self-studied English for 12.01 hours per week, among which the time on practicing English writing was 1.60 hours per week.

All the participants were randomly grouped into the control group (non-dynamic assessment group, NDA group) and the experimental group (dynamic assessment group, DA group). An independent-samples *t* test was conducted on the scores of Comprehensive English III and the results showed that the subjects between the NDA group ($M=79.05$, maximum possible=100) and DA group ($M=80.45$, maximum possible=100) were comparable in English proficiency ($t=.547$, $p=.563$). In the control group, the participants were given traditional instruction in writing, while in the experimental group, dynamic assessments were adopted to provide graduated mediation for individuals. More details are discussed in the mediation section.

Instruments and Measurement

This experimental study, to measure the effectiveness of dynamic assessment with support from online scoring systems in English writing class, employed three tests: the pretest, immediate posttest and transfer posttest. The purposes of the pretest were to ensure the comparability in English language proficiency in the two groups and to compare with the two posttests afterward to measure the improvement in English writing. An immediate posttest followed the mediation part to measure language learner's improvement right after different treatments. And a transfer

posttest was designed after ten days after the immediate posttest to scale participants' improvement in the long term. All three tests were conducted in the same form: a writing task on popular topics in society to date. Participants were required to finish the writing task in 40 minutes independently. However, because of the outbreak of new coronavirus, participants were not gathered together. Rather, they stayed in their places and finished the writing tasks without supervision from any teachers or researchers. Besides the time limit, they were also informed of the measurement of the writing, that is, grades were given according to the performance in coherence, complexity, accuracy and structure. Thus, participants had no idea of what data the researchers focused on. After the time limit, participants sent their works to the researchers by email.

The writing was rated by Pigaiwang, Grammarly, and two teachers. Pigangwang, a very widely-used online writing scoring system in mainland China, can provide information and feedback in coherence, complexity, structure and accuracy by matching the utterances in corpora. However, according to He (2013), the online automatic scoring system enjoyed high reliability ($t=-2.8$, $p=.008$) but the feedback was provided in line with sentences instead of the whole essay. To improve the quality of assessing writing, Grammarly was used as a supplemental tool because it can improve students' writing in spelling mistakes, identify fragments and provide suggestions on all categories of word form (Daniels & Leslie, 2013; Ghufon & Rosyida, 2018). Due to the deliberation on similar suggestions from the two online automatic scoring systems, two raters compared those outcomes to remove overlapping parts manually.

Because the two online scoring systems failed to assess writing in terms of coherence, complexity, and structure, and manual work in judging those aspects may cause subjectivity to different degrees, this paper only measured such linguistic errors as subject-verb agreement, article error (overuse, misuse and omission), verb forms, sentence structure, conjunction error, adjective and adverb misuse, preposition error, and the like. Participants were given 50 marks. One point shall be deducted from the total score for one linguistic error until all points were deducted.

A questionnaire and interview were also employed to facilitate data collection. After all tests, a questionnaire was delivered to the DA group to investigate the difficulties they had encountered in the prompts and their preference for DA and 4 interviewees from the DA group (two with a large ZPD and two with a small ZPD) were selected randomly to describe how they finished their writing and revisions at different phases.

Mediation

A graduated prompt approach (Brown & Ferrara, 1985; Campione & Brown, 1990), in which prompts are given from the implicit to the explicit, was adopted in the experimental group.

At the initial phase, the teachers input students' works into Pigaiwang and Grammarly to obtain useful advice on accuracy and provided overall feedback on the coherence, accuracy, complexity and structure for participants' reference. What should be noted is that the mediation

was embedded in the raters' feedback. And three prompts were generated from an implicit way to an explicit way. For example, instance (1) is the comment form in the first prompt:

(1) The writer has used some cohesive devices to make the essay smooth (coherence). The writer has a good foundation in diction and most advanced vocabulary has been properly used. And the author has also employed some complex long sentences, but more complex long sentences are encouraged to obtain a higher score (complexity). Such grammatical errors as subject-verb agreement, conditional clause, tense occur here and there (accuracy). The structure is rather clear (structure).

The teacher returned the first drafts with general feedback to participants and asked them to revise their writing in 30 minutes according to the overall comments given. After revision, they handed in the second drafts while the teacher fed back again. But in the second drafts, with the facilitation of the writing webs, the second prompt underlined the locations of the errors to call for further revisions, as shown in (2).

Check tense and logic

(2) When some people are enjoying eating wild animals, a new terrible disease called covid-19 emerges and spreads widely. With the ever-expanding of coronavirus area, more than 100 countries has involved and taken actions.

Check preposition, singular and plural forms, and subject-verb agreement

Again, the researchers sent the second drafts back to participants with feedback and permitted 30 minutes to produce and submit the third drafts in line with the second prompts given. In the last phase, the writing webs would assess and correct all the errors in the third drafts, presented in (3). After the third mediation, participants were given 30 minutes to read to understand the revisions, then the writing task was finished hereby.

The control group, meanwhile, received no graduated prompt, but explicit feedback after the assessment of each writing task. That is, teachers provided correct forms directly for their linguistic errors without further explanations, as shown in (3).

Because

(3) When some people eat wild animals, a new terrible disease called covid-19 emerges and spreads widely. In the ever-expanding of coronavirus areas, more than 100 countries has involved and taken actions.

have

Procedures

After 44 participants' consent on attending the writing research, the researchers divided those subjects into two groups randomly: the NDA group DA group. Then, each group followed four steps to accomplish the experiment as displayed in Figure 1.

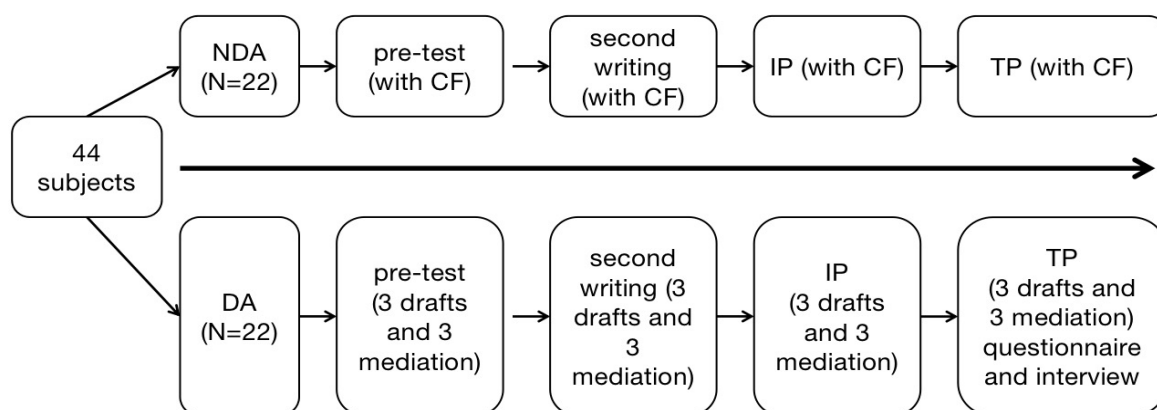


Figure 1. *Research procedures*

In the teaching experiment, participants finished four writing tasks: each step covered one writing task. The first step was the pretest, the third the immediate posttest (IP), the fourth the transfer posttest (TP). The experiment lasted for 3 weeks, with pretest arranged in the first day, second writing on the fourth day, immediate posttest on the seventh day, and transfer posttest on the 18th day.

The difference between the two groups was the assessments. In the NDA group, students' writings were assessed traditionally. Participants were provided timely corrective feedback (CF) according to the coherence, complexity, accuracy and structure in writing. Meanwhile, they were not expected to revise their writing in line with the feedback provided. The procedures in the following three sessions were identical. Subjects were given one score per writing task.

In the DA group, participants, like those in the NDA group, finished their writing in the first hand. Then, the teachers would provide mediation for participants to revise the works in 30 minutes. Altogether, they received three prompts and submitted three drafts on the same topic (40 minutes for the first draft, 30 minutes for the second and the third draft). Moreover, they were also given three marks in the same writing. The second writing, IP and TP sessions have repeated the procedure in the pretest.

Results

Skewness and Kurtosis have scrutinized all the data and they were all normally distributed. Hence, the data were processed by parametric tests with the alpha value setting at .05. An independent-samples *t*-test was run on pretest to reassure the equality in writing accuracy between groups and the data showed that there was no remarkable significance between the

NDA group and DA group ($t=-.342$, $p=.734$). In other words, at the onset of the experiment, NDA and DA group were similar in the accuracy in writing.

Paired-samples t -tests were conducted to provide an overall description for all participants in the three tests and the results were shown in Table 1. Compared with the results in the pretest, participants have made impressive improvements in both IP ($t=-5.887$, $p<.001$, $d=0.889$) and TP ($t=-5.226$, $p<.001$, $d=0.789$). However, no statistical significance was found between IP and TP ($t=.573$, $p>.05$) and a marginal regression in accuracy was observed in TP. In the writing, article omission, noun and noun phrase, subject-verb agreement, preposition errors were more prevailing than such errors as adjective errors, conjunction misusing.

Table 1.

Overall comparisons in improvement

		N	Mean	SD	t	Sig. (2-tailed)
Pair 1	pre	44	32.86	5.378	-5.887	.000*
	IP	44	37.64			
Pair 2	pre	44	32.86	5.740	-5.226	.000*
	TP	44	37.39			
Pair 3	IP	44	37.64	2.894	.573	.570
	TP	44	37.39			

[note]: * refers to Sig. $<.05$.

In this part, the findings on the three tests and post-questionnaire are presented to answer the research questions.

Within-group Comparisons

Operating paired-samples t -tests, within-group comparisons in the NDA group and DA group were made respectively, as results were shown in Table 2. To observe the NDA group's improvement in accuracy in writing, the pretest was compared with IP and with DP, respectively and IP was compared with TP by running paired-sample t -tests. The results exposed that the participants had a marginal higher accuracy in IP ($t=-1.854$, $p>0.05$) and in TP ($t=-1.415$, $p>0.05$) than in the pretest. The regression between IP and DP can be ignored ($t=0.781$, $p>0.05$). In other words, subjects in the NDA group can not see statistically significant improvement in linguistic accuracy.

Table 2.

Within-group comparisons

		Mean		t		Sig. (2-tailed)	
		NDA	DA	NDA	DA	NDA	DA
Pair 1	pre	33.18	32.55	-1.854	-8.279	.078	.000*
	IP	35.09	40.18				
Pair 2	pre	33.18	32.55	-1.415	-7.048	.172	.000*
	TP	34.55	40.23				
Pair 3	IP	35.09	40.18	0.781	-.085	.444	.933
	TP	34.55	40.23				

[note]: * refers to Sig. <.05.

In the within-group comparisons of the DA group, participants had made a remarkable improvement in IP, compared with the pretest ($t=-8.279$, $p<0.01$). Since the statistical significance between the two tests, Cohen's d was calculated and found that the effect size was large ($d=1.763$). The same situation occurred in the comparison between pretest and TP ($t=-7.048$, $p<0.01$, $d=1.503$). To be more specific, both IP ($p<0.001$) and TP ($p<0.001$) witnessed a more noteworthy enhancement than the pretest. However, a very marginal retrogress was made between IP and TP ($t=-.085$, $p>0.05$). It could be seen that the remarkable improvement in the overall description was mainly caused by the DA group rather than the NDA group.

Between-group Comparisons

Between-group comparisons were implemented to measure the difference in NDA and DA group improvement. To minimize the effect of the group difference in pretest on the comparison, the researcher calculated three gains for each group: Gain 1 = IP - pretest; Gain 2 = TP - pretest; and Gain 3 = TP - IP. Independent-samples t -tests were run to reckon the difference between groups, and the outcomes were exhibited in Table 3.

Table 3.

Between-group comparisons

	Group	Mean	SD	t	Sig. (2-tailed)
Gain 1	NDA	1.91	4.830	4.143	.000*
	DA	7.64	4.327		
Gain 2	NDA	1.36	4.520	4.343	.000*
	DA	7.68	5.112		
Gain 3	NDA	-.55	3.277	.673	.505
	DA	.05	2.497		

[note]: * refers to Sig. <.05.

In Gain 1, both NDA and DA groups have obtained improvement in linguistic accuracy, but the difference in improvement was striking ($t=4.143$, $p<0.001$, $d=1.25$). In other words, both the DA group and the NDA group saw improvement in linguistic accuracy. However, participants in the DA group progressed significantly faster than those in the NDA group. Similarly, the same situation can be observed in Gain 2 ($t=4.343$, $p<0.001$, $d=1.30$). However, in the third gain, different trends occurred. Participants in the NDA group gained a higher linguistic accuracy in IP rather than TP without statistical significance. It meant that they experienced a slight retrogress, and the effects of the treatment faded with time passing by. Meanwhile, the subjects in the DA group underwent an improvement from IP to TP in linguistic accuracy, though not impressive ($p>0.05$). Thus, the effect of DA sustained through time and influenced participants' learning outcomes in the long term.

ZPD And Future Performance

ZPD is the gap between the first score in pretest before mediation and the third score after mediation. The Pearson r correlation coefficient was used to analyze the relationship between ZPD and Gain 1 ($p=.003$, $r=.610$), between ZPD and Gain 2 ($p<.001$, $r=.774$), and between ZPD and Gain 3 ($p=.012$, $r=.527$) and a significant correlation was found. To be more specific, the correlations between ZPD and Gain 1, and between ZPD and Gain 2 were substantial since the value of r is between .80 and .60, while the one between ZPD and Gain 3 was moderate with r value being the scope from .40 to .60.

Questionnaire

In the questionnaire delivered to DA participants, 100 percent of students reported their improvement in writing, including structure, content and accuracy and coherence. The result agreed with the quantitative findings in IP and TP.

In the mediation, a vast majority of students proposed that they would think and explore when the teacher firstly pointed out the types of errors and made suggestions on the revision. But, nearly half of the students failed to correct the errors in the essay because of the time limitation and ambiguity in locations and suggestions. It was not until the teacher pointed out the specific locations of the errors did some students realize them. Though explicit feedback worked more effectively than the implicit feedback in the short term, the implicit prompts had a better outcome in motivating them to think and revise. Nevertheless, they suggested locating the errors would be more effective.

When asked their preference for the mode of writing training, all subjects were favorable to the DA mode mainly due to four reasons: the major reason lied in that the graduated prompts promoted higher attention and stronger memory in the errors; another reason was that DA mode provided opportunities for them to conduct self-reflection and independent thinking; it also helped trainees to formulate a good habit in revising their essays; students can obtain a sense of satisfaction from the training. However, in the questionnaire, students also reflected some concerns on the time-consuming feature of this mode.

Interview

Four students (two with large ZPD and two with small ZPD) were invited to have a one-to-one online interview with the researcher to investigate how they finished the writing task in DA mode. For the two participants with large ZPD, they reported a positive response to the prompts during the writing process. In the first draft, they said that the time was very limited because they had spent much time in planning and compiling. Thus they hardly had any time to revise the grammatical errors in the draft, though they realized them.

When the first mediation was turned back to them, they reported some difficulties. Two interviewees with small ZPD and one with large ZPD reflected that the feedback was too ambiguous to understand since the teacher only pointed out there were mistakes in the essay without locating them. Hence, they used a lot of time to guess where were the errors. The other

interviewee with large ZPD told the interviewer that the feedback was not too hard to follow because he was strong in English grammar. The ways they handled those comments were different too. For those with large ZPD, they thought hard on the comments because they considered the corrections and revisions in language was essential in writing development. When coming across difficulties, they sought different ways to solve those problems. For example, they searched answers from the internet, turned to classmates/teachers for help, looked up dictionaries. They possessed great motivation in revising possible errors in their writings, too. However, the ways for those interviewees with small ZPD to revise their writings were quite different. They told the researcher that they tried to revise the errors but they did not know whether they were correct or not, so they provided a possible correction and left the revision unchecked. They thought that the teacher may point them out in the next mediation and there was no need for them to try so hard.

In the second mediation, the researcher pointed out and located their errors, so all the interviewees articulated the clearness and effectiveness of the prompts. Nonetheless, there were still differences in solutions. The participants with large ZPD tried to revise those errors and they asked themselves why those errors happened. They turned to dictionaries to see more examples to ensure their full understanding of linguistic rules. One of them stated that she failed to understand the researcher's comment so she raised her different opinions in the next draft and the teacher's responses helped a lot. But, as for those students with small ZPD, they revised their writings according to the teacher's feedback but did not seek for more information. For example, when the teacher pointed out the errors in subject-verb disagreement, they can easily respond to the comments, because there were only two forms: single and plural. They revised those errors without knowing why.

The third prompts revealed the correct forms for participants explicitly. Those subjects with large ZPD read the correct forms, revised their writing, and sought more information from different sources. All in all, they took the whole writing task as an adventure or as a game. For those students who had a small ZPD, they behaved differently again. They only read the feedback provider's suggestions but they did not try to revise them for the teacher did not ask for a revision draft.

When being asked what made a good essay, they all responded that a good composition should be logical in content and organized in structure. Using a lot of cohesive devices should be considered as good too. And only three participants (two with large ZPD and one with small ZPD) thought that grammar in writing was important too. The researcher asked the student with small ZPD why she did not seek more information to revise her writing, and she responded that she did not want to waste time in seeking for help because she had a bad memory to remember the correct forms. She would make the same mistakes in a new piece of writing. The other three interviewees opined that they were attentive to the errors they had made in the previous writing to reduce errors when working on a new piece of writing. From the interview, attention, critical thinking, engagement and motivations are important factors in influencing the effectiveness of mediation.

Discussion

The Superiority of DA

In terms of the first research question, the NDA group has improved in the experiment, but the improvement was at a slow pace and a regression could be seen with time passing by. And as for the DA group, participants improved their accuracy remarkably in the immediate posttest, but the pace of progress slowed down when the DA mode in writing ceased. Between-group comparisons were implemented to answer the second question: participants in the DA group outperformed marvelously NDA group in both IP and TP. That is, in the short term, DA can much more effectively improve learners' accuracy in writing, and in the long term, the learning outcome can still be seen. The findings are consistent with those of Xiaoxiao and Yan (2010), Hassaskha and Haghparast (2012), Davoudi and Ataie-Tabar (2015), and Derakhshi (2019). The reasons why DA is more effective than traditional assessments are probably explained SCT (Vygotsky, 1978). Mediation and interaction are two important elements in SCT (Watson-Gegeo, 2004) and DA provides prompts through interactions.

In the DA treatment, the teacher interacts with the language learners literally and dynamically. Therefore, the information acquired from the process of critical thinking and engagement is collected and the new cognition is intermentalized (Lantolf & Poehner, 2007). In the meantime, interactions and mediation can raise language learners' notice and attention on certain linguistic forms, especially those students who view linguistic accuracy as an important element in making a good essay. Notice and attention are the onsets of learning and acquisition (Mackey, 2006; Robinson, 1995). This assumption has been verified in the exploration in the interview after the treatment that students were alert to reduce the previous errors in a new piece of writing. That might be the reason why participants can improve their linguistic accuracy in the short term. Besides, since the mediation provided in DA is graduated, language learners need to conduct critical thinking and seek correct forms from different resources, when the mediation is implicit (Zarbaftian et al., 2020). The implicit mediation also results in a high engagement in revisions (Vakili & Ebadi, 2019). High engagement and critical thinking are helpful in intramentalizing new knowledge (Lantolf & Poehner, 2007). Thus, in a new piece of writing, the language learners can solve the linguistic errors independently. Another possible explanation is the explicit and implicit input. In the NDA group, the mediation was explicit and no revision was required afterward. Though language learners may hold a strong awareness of such rules, the input is still declarative and learners can not transform it into implicit, because explicit and implicit knowledge is processed in two separate systems (Krashen, 1981). However, DA adopts graduated mediation so it is successful in maintaining the advantages from implicit knowledge and explicit knowledge. In the first and second mediation, language learners should conduct critical thinking to deal with the implicit knowledge provided. In contrast, in the last mediation, explicit knowledge makes up those aspects that the implicit knowledge does not work. In this way, language learners can access correct forms to improve their linguistic accuracy.

The Predictive Power of DA

The result of the third research question provides statistical evidence to Lantolf and Poehner's (2007) proposal that ZPD helps teachers to understand students' future learning outcomes better. ZPD's capacity in predicting future performance is probably due to its reflecting participants' motivation and their efforts in learning. The interaction or the guidance works as a tool to encourage students to internalize those explicit knowledge presented by thinking independently. Moreover, for those language learners with high motivation, they seek other resources for help and process the information unconsciously. The input can serve as implicit knowledge. The two kinds of information are more likely to arm language learners with comprehensible input to work out a solution to the problem in a new piece of writing. Therefore, a larger ZPD probably means more critical thinking and more input, and it may lead to better learning outcomes.

Conclusion

This paper has employed online scoring systems to simplify the process of feeding back under the framework of interventionist DA. It has provided evidence for its positive role in promoting learning efficiency. In the meantime, to some extent, it has suggested a possible way to reduce writing teachers' workload in instruction to implement DA in a large-scale classroom.

However, this paper is not free from limitations. First, it is hard to determine the reliability of the writing outcomes. Due to the widespread of COVID-19, participants had online classes at home, and they were not gathered to finish the writing tasks in the time given. Instead, they were only informed of the requirements and finished their writing with their self-regulation. So, it is hard to judge whether the participants finish the task in a limited time or whether they compiled the essay independently. Second, linguistic accuracy should not be the only criteria to evaluate the quality of an essay. This paper only investigated linguistic accuracy because online scoring systems are reliable and more valid in accuracy than in coherence, content and structure. Third, individual differences are excluded from the present study. From the qualitative data, those participants who are highly motivated tend to gain more benefits from DA treatment. But due to the space limitation, this paper does not identify its mediating role in improving linguistic accuracy.

References

- Ableeva, R., & Lantolf, J. (2011). Mediated dialogue and the microgenesis of second language listening comprehension. *Assessment in Education: Principles, Policy & Practice*, 18(2), 133-149.
- Afshari, H., Amirian, Z., & Tavakoli, M. (2020). Applying group dynamic assessment procedures to support EFL writing development: Learner achievement, learners' and teachers' perceptions. *Journal of Writing Research*, 11(3).
- Ahmadi, A., & Besharati, F. (2017). Web-based Versus Face-to-Face Interactionist Dynamic Assessment in Essay Writing Classrooms—A Comparative Study. *The Journal of Language Learning and Teaching*, 7(1), 1-29.
- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301-320.

- Allal, L., & Ducrey, G. P. (2000). Assessment of—or in—the zone of proximal development. *Learning and Instruction*, 10(2), 137-152.
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576-598.
- Brown, A., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 273–305). Cambridge: Cambridge University Press.
- Campione, J. C., & Brown, A.L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141–172). Hillsdale, NJ: Lawrence Erlbaum.
- Daniels, P., & Leslie, D. (2013). Grammar software ready for EFL writers. *OnCue Journal*, 9(4), 391-401.
- Davoudi, M., & Ataie-Tabar, M. (2015). The effect of computerized dynamic assessment of L2 writing on Iranian EFL learners' writing development. *International Journal of Linguistics and Communication*, 3(2), 176-186.
- Derakhshi, Z. (2019). On the Role of Dynamic Assessment on Promotion of Writing Linguistic Accuracy among EFL Learners: An Interventionist Model. *International Journal of Research in English Education*, 4(2), 14-28.
- Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, 12(4), 395-403.
- Guterman, E. (2002). Toward dynamic assessment of reading: applying metacognitive awareness guidance to reading assessment tasks. *Journal of Research in Reading*, 25(3), 283-298.
- Hassaskhah, J., & Haghparsat, M. J. (2012). A comparative study of the impact of DA models on the writing ability and attitude of Iranian EFL learners. *The Buckingham Journal of Language and Linguistics*, 5, 38-51.
- He, Xu-liang. (2013). Reliability and validity of the assessment by the Pigaiwang on college students' writings. *Modern Educational Technology*, 23(5), 64-67.
- Hidri, S. (2019). Static vs. dynamic assessment of students' writing exams: a comparison of two assessment modes. *International Multilingual Research Journal*, 13(4), 239-256.
- Lantolf, J. P. (2000). *Sociocultural theory and second language learning* (Vol. 78, No. 4). Oxford: Oxford University Press.
- Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching*, 42(3), 355-368.
- Lantolf, J. P., & Poehner, M. E. (2007). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics and Professional Practice*, 1(1), 49-72.
- Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). Sociocultural theory and second language development. In B. VanPatten and J. Williams(Eds.), *Theories in second language acquisition: An introduction* (pp. 207-226). New York: Routledge.
- Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. Guilford Press.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-430.
- Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2019). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 1-30.
- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, 43(3), 471-491.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research*, 17(3), 323-342.

- Poehner, M. E., & Leontjev, D. (2020). To correct or to cooperate: Mediational processes and L2 development. *Language Teaching Research*, 24(3), 295-316.
- Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337-357.
- Robinson, P. (1995). Attention, memory, and the “noticing” hypothesis. *Language Learning*, 45(2), 283-331.
- Shabani, K. (2018). Group dynamic assessment of L2 learners’ writing abilities. *Iranian Journal of Language Teaching Research*, 6(1), 129-149.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Vakili, S., & Ebadi, S. (2019). Exploring EFL learners’ developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning*, 1-36.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press
- Watson & Gegeo, K. A. (2004). Mind, language, and epistemology: Toward a language socialization paradigm for SLA. *The Modern Language Journal*, 88(3), 331-350.
- Xiaoxiao, L., & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33(1), 24-40.
- Yang, Y., & Qian, D. D. (2019). Promoting L2 English learners’ reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, 1-25.
- Zhang, Y. (2013). The theoretical construction of dynamic assessment mode in Chinese tertiary EFL writing class with online teaching and scoring system. *CALL-EJ*, 14(2), 38-50.
- Zhang, J., & Lu, X. (2019). Measuring and supporting second language development using computerized dynamic assessment. *Language and Sociocultural Theory*, 6(1), 92-115.
- Zarbaftian, A., Abbasian, G. R., Mohseni, A., & Baradaran, A. (2020). Synthesizing transformative education with dynamic assessment in developing EFL learners’ productive skills. *Cogent Arts & Humanities*, 7(1), 1-26.

Acknowledgments

Not applicable.

Funding

Not applicable.

Ethics Declarations**Competing Interests**

No, there are no conflicting interests.

Rights and Permissions**Open Access**

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. You may view a copy of Creative Commons Attribution 4.0 International License here: <http://creativecommons.org/licenses/by/4.0/>.