Original Article

MALL-based Cumulative Group Dynamic Assessment in Remote Teaching: The Case of Distance Education in Iran

Amir Ghajarieh *[1], Mohammad Amin Mozaheb [2] ،Samira Mohajer[3]

1. Assistant Professor, Department of TEFL, University of Ershad Damavand
2. Assistant Professor in TEFL, Language Center, Imam Sadiq University, Tehran, Iran
3. M.A. University of Ershad Damavand

Abstract

While various aspects of Group Dynamic Assessment (GDA) have been explored in previous studies, cumulative GDA in distance education is an under-researched area. This study aims to examine the effect of teaching cumulative GDA in English as a Foreign Language (EFL) settings with a focus on discourse markers. To this end, a mixed-methods design was adopted with speaking tasks of First Certificate in English (FCE) test and interviews used as the main instruments for collecting data in the quantitative and qualitative phases of this study. A total of 64 homogenized female EFL learners were randomly divided into experimental and control groups. Cumulative GDA instruction of discourse markers in the experimental group was conducted by a teacher well-trained in intervention workshops. The analysis of the data based on t-test and analysis of covariance (ANCOVA) showed a positive effect of GDA on the test group. In the qualitative phase of the study, interviews were used as the tool to reveal the attitudes of ten participants in the experimental group who opted to attend interview sessions. The content analysis of the interview results revealed that they had varied attitudes towards online cumulative GDA, and dominantly favoured this teaching approach in distance learning. The results highlight the importance of Assessment for learning (AfL) over Assessment of Learning (AoL) in distance education. Findings have implications for policy, practice and research.

Keywords

Group Dynamic Assessment (GDA), Online GDA, EFL learners, Speaking abilities, Distance education.

Introduction

A growing interest has been directed toward the issue of integrating assessment with teaching and learning through the lens of applied linguistics (Beaumont et al., 2011; Poehner, 2009). Assessment for learning (AfL) over Assessment of Learning (AoL) has been focused on recent studies with a change of lesson presentation in EFL/ESL/ESP classes from the conventional Presentation, Practice Productions to Testing Teaching and Testing, as noted by distinguished scholar Scott Thornbury. In this context, Dynamic Assessment (DA) has invited considerable L2 research attention. Such a practical approach to teaching and testing could be of significant value for distance education with blended teaching and flipped learning as two student-centred approaches in pedagogy (So et al., 2021). Feedback is a difficult task to accomplish in distance education as it often lacks the traditional in-person instruction sessions, Nevertheless, DA with a unique blended approach to teaching and testing that brings guidance and feedback for educators (Poehner, 2009; Ouden, 2020) could arguably increase the efficacy of distance learning. DA provides individualized or group instruction and feedback to assist learners in deepening their understanding through low-stake feedback and consecutive assessment. While traditional physical classes may resist new methods in teaching students, the very nature of

distance education could be more welcoming of such vanguard change in teaching and learning.

In other terms, DA can serve as a powerful tool for distance learning where students are encouraged to critically think of their learning process and seek more feedback from their teachers and peers. As such, DA can potentially maximize the effectiveness of distance education. Most research in mainstream education has focused on various dimensions of DA (e.g., Abdulaal, 2022; Rashidi, & Bahadori Nejad, 2018). Nevertheless, there is a paucity of data on how such an approach of blended teaching and assessment could be effective in distance education.

Dynamic Assessment is a key concept of Vygotsky's Sociocultural Theory (SCT) and relying on Vygotsky's works, Poehner (2005) integrated testing and instruction. Poehner (2009) introduced the term Group Dynamic Assessment (GDA) This concept is based on the guideline that gradual feedback from the teachers at the social level can support the ZPD of learners which boosts their competence as well. The efficiency of GDA in improving different skills and aspects of language has been depicted in previous studies (see, e.g., Miri, et al., 2017; Farahani, & Moghadam, 2018). One kind of GDA is cumulative GDA in which students communicate with their teachers one by one. According to Poehner (2009), when a student produces an erroneous utterance, the teacher offers mediation for the learners to support the learners to overcome their problems.

DA argues that assessment should be accompanied by instruction (Lidz & Gindis, 2003). That is to say, as noted by Beaumont et al. (2011), assessment and instruction are interrelated given that assessment contributes to teaching and learning. Poehner (2009) believes that DA can pave the way for language teachers to create meaningful interactions with students by providing appropriate mediation. Cumulative Group Dynamic Assessment may ensure more interaction and feedback which are remarkably needed in distance education due to far less face-to-face contact between the educator and their students.

Most research in mainstream education has focused on various dimensions of DA (e.g., Abdulaal, 2022; Rashidi, & Bahadori Nejad, 2018). Nevertheless, there is a paucity of data on how such an approach of blended teaching and assessment could be effective in distance education. Particularly researchers in language education have been largely interested in the application of DA over recent years. Although studies have proved the efficiency of GDA in improving various skills of language, including writing accuracy (Tabatabaee, et al., 2018) or collocations (Farahani, & Moghadam, 2018, DA has not sufficiently been studied in terms of improving or enhancing L2 learners' speaking skills except for a few exceptions, including Son and Kim (2017) in the Korean educational context. More specifically, there has also been a scarcity of research, particularly on discourse markers through cumulative group dynamic assessment. Thus, the current study fills this gap and explores the potential efficiency of online cumulative GDA in improving Iranian EFL learners' speaking abilities with a focus on discourse markers in distance education. This empirical evidence can shed more light on the under-researched interplay of distance education and language learning. Additionally,

For the purposes of this study, three research questions were formulated:

RQ1: Does online MALL-based cumulative GDA significantly enhance the speaking abilities of Iranian EFL students?

RQ2: To what extent, can online MALL-based cumulative GDA with discourse markers in focus improve Iranian EFL learners' speaking abilities?

RQ3: What are Iranian EFL learners' attitudes towards online MALL-based cumulative GDA conducted in this study?

Review of the Literature
Reform in language teaching approaches has always affected language assessment. The earlier

L2 teaching methods were language-centered or learner-centered and used a product-oriented syllabus. Later, language teaching methods including natural and communicational approaches became learning-centered and moved toward process-oriented approaches where the syllabus included communicative tasks focusing on classroom interaction rather than linguistic features (Kumaravadivelu, 2006).   Influenced by changes in language teaching, the pendulum of language assessment shifted from a product-oriented methodology to a process-oriented one (McNamara & Roever, 2006). Emphasizing the social dimension of language testing, McNamara (2000) states that language assessment does not merely involve a single test under time constraints to evaluate the learner's performance but learners' performance during the course is taken into consideration.

Assessment of students' learning is important in language learning and teaching. Static assessment methods, which include standardized calculation and norm-referenced tools, are typically used in the process. Static assessment fails to mention details about the learning and teaching processes and provides less authentic tasks (Nazari & Mansouri, 2014). Contrary to static assessment, the concept of DA integrates teaching and assessment and focuses on the process and improvement of learning (Poehener, 2005, 2009). DA, primarily a Vygotskian idea, penetrated different fields of study including the L2 domain to challenge the traditional methods of assessment implemented by teachers, researchers, and curriculum designers. Collaborative engagement is the intermental functioning that should be occurred through the process of an individual development to finally result in the intramental functioning or functioning independently and without the need for external support. Moreover, to reach the intramental functioning stage, learning and teaching should be done by observing the Zone of Proximal Development (ZDP), providing mediation to help and support learners' development of abilities (Poehner, 2009).

Teaching and testing are not two separate activities and are regarded as interconnected (Poehener, 2005). Relying on DA principles, Poehner (2009) highlights the significance of interaction and social mediation within the context of the classroom under the framework of GDA. Accordingly, teachers act as the mediator who negotiates and cooperate with a group of learners leading them forward in their ZDP.

Based on GDA, the feedback by the expert/assessor/ teacher to the learner is given through dialogue and at the social level, and other learners present in that dialogic context can make use of the feedback exchanged or they can participate in feedback exchange to co-form further awareness (Alemi et al., 2019). In fact, GDA follows the same principles of DA in offering mediation to assist students to co-construct a ZPD, which is not an individual one but a group-based one.

Concurrent vs. cumulative models of dynamic assessment.
Poehner (2009) put forth concurrent and cumulative as two different approaches to GDA. The second of which was the current study's focus. The concurrent approach refers to the GDA activity in which the instructor, as the primary interactant, communicates with the whole community of students. Moreover, in some cases, the teacher may offer mediation to individuals as secondary interactants. Nevertheless, there is a quick shift between the primary and secondary interactants since a challenge by an individual provides the condition to contribute to the learning of another individual in the group. Therefore, this model of GDA includes development within individuals' and group ZDPs concurrently.

In the cumulative GDA, each person engages directly in an activity, interpreting that each one-on-one interaction builds based on earlier interactions observed and witnessed by the entire class (Poehner, 2009).

Spoken discourse markers

Discourse Markers (DMs) are responsible for the coherence of discourse. Aijmer (2002) suggested that DMs cannot be limited to one part of speech and also they can be subcategorized based on "textual" and "interpersonal" levels (p. 25). Extending the work of Aijmer, Fung and Carter (2007) suggested interpersonal, referential, structural, and cognitive categories based on function. Biber et al. (1999) also classified types of DMs as interjections, greeting and farewell expressions, stance adverbials, response forms, vocatives, various polite speech-act formulae, and hesitators.

In sum, DMs can be defined as an important component with different grammatical classes, and functions to keep the listener and speaker focus on the ongoing discourse.

Studies conducted on GDA
Köroğlu (2019) studied DA for the evaluation of enhancing student teachers' speaking abilities and focused on their attitudes toward complex evaluation. The study's participants included 29 student teachers studying in the English language teaching department of a Turkish state university. The study's findings revealed that the participants held a positive viewpoint toward this appraisal style. As a result, it was recommended that DA can be used to assess speaking. Eventually, it was determined that the interventionist model maintained and improved student teachers' oral skills.

Yakışık and Çakır (2017) investigated the impact of complex evaluation on developing the speaking skills of ELT learners at a state university in their research. At the end of the experiment, the reactions of the students to the assessment form were analyzed using qualitative results. In transition experiments, the participants were found to be less dependent on the teacher's feedback, demonstrating the power of relationships in the students' ZPD. Besides that, findings from the student appraisal form revealed that students considered the testing technique to be helpful.

Poehner (2005) attempted to propose an alternative way of conceptualizing the interaction between evaluation and teaching in a doctoral study on the impact of DA on the oral proficiency of French L2 students. He attempted to investigate DA application to an L2 learning sense in this research. L2 DA was specifically considered in terms of the following: (1) some insights into learners' abilities that were not gained by evaluating a candidate's individual success but only became apparent during DA; and (2) the prospect of encouraging learners' abilities by DA. (3) the utility of DA in resulting in individualized instruction that remained vulnerable to examinees' ZPD; and (4) the degree to which progress transcended the given evaluation background. Six advanced undergraduate L2 French learners were asked to produce a sequence of oral narratives in French based on short video clips. The first narrative was produced directly by the students, and the second narrative was created with assistance from the examiner. The findings of these tests were used to create a six-week intervention course in which students met with the researcher for one-on-one teaching. These sessions aimed to resolve the issues found during the tests while taking into account their experience with the examiner during DA. Following enrichment, the type and scale of their improvement are investigated using repetitions of the initial tests as well as variants on all these activities. According to the results, DA was an important method of interpreting learners' skills and assisting them in overcoming linguistic problems. As a tool for making formative evaluation processes more comprehensive, the technique is particularly applicable in L2 classrooms.

Since the introduction of GDA by Poehner (2009), a good number of studies (e.g., Hidri, 2014; Miri et al., 2017) conducted in this regard in EFL/ESL context. Given the language proficiency, most of the studies have focused on EFL writing (e.g., Shrestha & Coffin, 2012) and listening comprehension (e.g., Farangi & Kheradmand Saadi, 2017). However, as the review of related literature showed, a few studies investigated speaking with a focus on DMs in this research line. The use of Mobile-Assisted Language Learning as a subcategory of CALL

(Computer Assisted Language Learning) also adds a new dimension to this research addressing a pressing need for Emergency Remote Teaching (ERT) and distance education in the post-pandemic era.

METHODOLOGY
Design of the Study
This research adopts a mixed-methods approach to collect quantitative-qualitative data using speaking tests and interviews. The research method for the quantitative phase of the study included an experimental study with cumulative group dynamic assessment as the treatment for the experimental group with t-test and ANCOVA run through IMB SPSS software version 27.0.1 as the statistical measures to examine the effect with quantitative data.   In the second phase, the quantitative data were obtained through interviews to find out the perception of students in the experimental group regarding the treatment. The qualitative data were analyzed through content analysis based on open coding, sub-themes, and overarching themes (Braun and Clarke, 2006; Maguire & Delahunt, 2017; Riazi, 2016) with emerging themes showing the interplay of dynamic assessment in distance learning. In the words of Creswell and Plano Clark's (2007, p. 68), "Embedded design" was used in this study given that the qualitative data analysis could enhance the overall design.

Participants
Ninety female English learners (aged 25 to 32) at B2 level, who studied at Atlas language academy in Tehran, were selected using convenience sampling. The FCE exam of Cambridge ESOL (2015) was then implemented to homogenize the sample's language proficiency. As a result, sixty four homogenous learners (31 EFL learners in the experimental group and 33 learners in the control group each comprising 3 classes) served as the participants. It should be noted that an EFL instructor taught both groups and was trained through intervention workshops on dynamic assessment and cumulative group dynamic assessment. for one month prior to implementing the treatment. The workshop was designed and guided by one of the authors who consulted with two experts in distance education and were quite familiar with such assessment. A total of ten participants from the experimental group also opted to be interviewed in the qualitative phase of the study. All the participants were told regarding the objectives of this research and granted their informed consent. The ethical release for the study was also obtained from the ethical committee of Ershad-Damavand University for conducting this research.

Instruments
For the purpose of the current study and to investigate the hypothesis, the following instrumentations were utilized:

First certificate in English(FCE) test for homogenization
Prior to conducting the treatment, the FCE test, which was chosen from a standard test bank by Cambridge ESOL (2015) test was applied to homogenize the participants. It should be noted that the FCE exam is planned and developed by Cambridge ESOL, and it shows language comprehension at level B2 of the Common European Framework of Reference (CEFR). FCE test has two parallel versions including computer-based version and paper-based version. It should be pointed out that the paper-based version was used in the current study due to its ease of administration.

   In this study, the writing module was excluded and the other three modules; namely, reading, listening and speaking, were administered among the considered population. The total score of the test was calculated by considering forty percent of the weight for the first module and twenty percent for the other modules of the test. It also should be mentioned that FCE test was

administered by the teacher of both control and experimental groups who had ten years of experience in teaching English language courses including FCE ones and was an M.A. holder in Teaching English as a Foreign Language (TEFL).

Further, administering the speaking module of the FCE test generally requires two examiners. One of the examiners asks a question and the other one listens and takes notes. As a result, another teacher having 15 years of experience teaching English and an M.A. degree in TEFL co-administered the FCE speaking section.

To determine the learners' speaking proficiency, one of the researchers utilized the speaking section of the FCE test used at the outset of the study. She used the speaking part of another edition of FCE (2010) as the speaking posttest at the end of the course, which lasted ten sessions, to assess the students' speaking ability during the treatment time.

Interview
In this study, using an interview, the participants' attitudes toward cumulative GDA were scrutinized to obtain a more comprehensive understanding of cumulative GDA. To maintain the participants' privacy and confidentiality, an Informed Consent Form was developed by the researchers based on Creswell's (2013) sample form. After gaining permission from the participants, the consent form was sent to each of them to be read and signed before the beginning of the interview sessions (Creswell, 2013).

After the treatment, only ten participants from the experimental group gave their consent to stay in the qualitative phase of the study and attended the interview sessions on a semi-structure basis. The interviews, which included five questions regarding learners' attitudes towards the use and appropriateness of GDA, is based on the one used in the study conducted by Tarighat and Khodabakhsh (2016) with different wording. Although the questions were taken from the abovementioned reference, for ensuring the content validity of the interview, the questions were checked and modified by one expert (A PhD holder of TEFL). The time allocated to the interview was about 20 minutes. Since all the students were at B2 level, the medium of the interview was English and it was conducted online via Skype.

Procedure
The experimental group attended online sessions by joining their Skype groups, which is considered as a type of Mobile-Assisted Language Learning (MALL) setting and were exposed to teaching and practicing speaking FCE sample tasks through GDA. The comparison group was taught without any guidelines of cumulative GDA. The whole course lasted 16 sessions, twice a week, and about 90 minutes. Ten sessions were allocated for teaching speaking ability through GDA and the rest was allocated to conducting the tests and interviews. The details of these sessions are as follows. In the first session of the course conducted in the experimental group, the speaking module of FCE was administered as the pre-test. The pretest was also held to diagnose the language learners' area of difficulty in speaking skills. The participants' performance on the speaking pre-test showed that their main area of difficulty in speaking ability was related to the use of discourse markers. So, the researchers decided to focus on this sub-skill of speaking ability namely discourse markers. Following the pre-test administration, both groups received ten sessions of instructions. It should be pointed out that the researcher just focused on the speaking tasks of FCE (parts 3 and 4) during the intervention procedure.

In the experimental group, learners were taught and assessed based on cumulative GDA following the guidelines proposed by Poehner (2009). Grounded on the suggestions given in the literature, particularly considering the studies conducted by Poehner (2005), Poehner (2009), and Aljaafreh and Lantolf (1994), the experimental group was taught as follows. Initially, the tasks taken from the FCE speaking tasks (parts 3 and 4) were given to the participants in each session.

The cumulative GDA interactions started with the teacher's selection of an issue related to the topic, then, the teacher initiates classroom talk with a student while encouraging other students to contribute to the spoken discourse around a topic. Particularly, when the learner produced an error in each session, the teacher provided her with the most implicit to the most explicit prompt to not only identify her area of difficulty but also to support her to overcome her errors. In other terms, upon failing to detect or correct errors on the part of students, the teacher goes on a continuum from the most explicit to the most implicit forms of feedback. For example, in case of asking the learner to repeat as a type of mediation, the Mediator usually required the learners to repeat what they have just said during the task completion. This type led the learners to reconsider what they were producing from linguistic or functional aspects. Below is a sample correction, the mediator requested the learner to overcome the error she had in using a discourse marker linguistically simply by asking her to repeat what she has just said.

  Sample 1
1.   Sarah : ... *in first picture , the woman
2.   who is telephoning someone is a clerk.
3.   M : sorry , could you please repeat what you just said ?
4.   in what ?
5.   Sarah : (...) in the first picture , the woman calling someone is a clerk.

Additionally, the discourse markers have been focused on in most sessions since the assessment of learning through the speaking pre-test showed that the learners had the most difficulty in using discourse markers. Another reason for focusing on discourse markers is that in the FCE test, discourse markers are taken into consideration by the assessors when the learner's performance on speaking skills is assessed. In the control group, the teacher did not implement cumulative GDA in the group. For the quantitative data, first descriptive statistics was adopted. For the inferential statistics, the researchers conducted an independent samples t-test and ANCOVA on the mean scores of the two groups.

At the end of the course, online interviews were conducted to gain qualitative data. For the second phase of the study, the recorded interviews with ten participants of the experimental group were transcribed verbatim based on Peraza's (2019) guidelines with a naturalized approach in transcription. The data was later content analyzed based on open coding, subthemes and overarching themes (Braun and Clarke, 2006; Maguire & Delahunt, 2017; Riazi, 2016). For qualitative analysis to be regarded as trustworthy, the researchers must take certain measures (Nowell et al., 2017). In this study, a list of codes and themes was identified by the researchers in the initial reading of the data and was later shared by an expert in coding qualitative data in education.

The revised codes were shared by two coders and were asked to content analyze 10 percent of the data. The Kappa coefficient for the inter-coder reliability was established to be 0.86, suggesting an acceptable agreement between coders. This measure was taken to ensure consistency in data analysis by two coders. During the content analysis of the data, new codes and emerging themes were notified to the research team and the coders implemented them in the final reading cycle of the data. Lastly, the frequency of each theme produced by the interview participants was calculated and anchor quotations from the participants were recorded for a detailed written report of the qualitative analysis.

RESULTS
Pre-test results

Table 1. illustrates the speaking pretest scores of the two groups.

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Experimental	31	12.00	18.00	15.2258	1.75042	-.190	.421
Control	33	11.50	18.50	14.8485	1.66060	-.022	.409
Valid N (listwise)	31						

According to Table 1, the mean scores of the two groups on the speaking pretest were 15.22 and 14.84. The inspection of skewness ratios of the two sets of scores (-.19/412 = -.46; -.022/.409 = -.05) showed that the distribution of both sets of scores was normal as they fell within the legitimate range for normal distributions, i.e., ±1.96. Therefore, to compare the pretest scores of the two groups a parametric independent samples t-test was run. The results of the independent samples t-test ($t_{(62)}$ = .885, p = .38 > .05) showed that there was no significant difference between the two groups in terms of speaking proficiency before the treatment was initiated.

Post-test results
Table 2 displays speaking post-test scores by the two groups.

Table 2. Descriptive Statistics of the Speaking Posttest Scores of the Two Groups

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Experimental	31	15.00	20.50	18.12	-.370	.421
Control	33	12.00	20.00	16.28	-.025	.409
Valid N (listwise)	31					

As illustrated in Table 2, the mean scores of the two groups for the posttest were 18.12 and 16.28. The skewness values of the two sets of scores (-.37/412 = -.89; -.025/.409 = -.06) showed that the distribution of both sets of scores was normal as they fell within the acceptable range for the normal distribution, i.e., ±1.96. Therefore, to compare the post-test scores of the two groups a parametric independent samples t-test was run. If the results show a significant difference between the two groups, this means that the difference between the control and experimental groups are statistically significant.

Table 3. Independent Samples t-test on the speaking posttest scores of the two groups

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	T	Df	Sig. (2 tailed)	Mean Difference	Std. Error Difference		Lower	Upper
Pretest	Equal variances assumed	3.715	.059	3.843	62	.000	1.84115	.47912		.88340	2.79890
	Equal variances not assumed			3.878	59.039	.000	1.84115	.47481		.89108	2.79122

The results of Levene's test of equality of variance (Table 3, first row) was non-significant (F = 3.715, p = .059 > .05). Therefore, the output is reported here with equality of variances assumed and a major assumption of ANCOVA is ensured. Based on the t-test results, the t-observed with the p-value less than 0.05 level of significance (t (62) = 3.84, p = .000 < .05, Cohen's d = .98, representing a large effect size), showed that there was a significant difference between the groups. In fact, a significant difference existed between the pretest and posttest scores of experimental and control groups in terms of speaking proficiency.

To ensure the difference was not related to pre-test scores and eliminate its effect, an analysis of covariance (ANCOVA) was run to control the variable as the covariate for a fair comparison

Table 4. Results of ANCOVA for comparison between the experimental and control groups' post test

	Sum of Squares	df	Mean Square	F	p	22
Overall model	150.384	2	75.192	304.035	<.01	
groups	150.174	1	150.174	595.474	<.01	0.90
Pre-test * groups	0.210	1	0.210	0.833	0.55	0.00
Residuals	15.384	61	0.252			

The analysis of covariance showed that the pre-test effect was not observed given that the F-value for the pre-test and groups interaction was .833 with a significance level more than .05. (p= .36) This indicates that the post-test results are not significantly affected by the pre-test results.

The results of the paired samples t-test and ANCOVA indicated that there was a significant difference between the experimental and control group's mean scores on the posttest. Thus, the null hypothesis is rejected which means teaching speaking through Cumulative GDA had a positive effect on the experimental group.

Discourse Markers in Group Dynamic Assessment

Due to the lack of awareness regarding the importance of discourse markers in spoken English and FEC as well as other proficiency tests, the teacher talked about the functions of such discourse markers in speech and daily conversations. Before the treatment, the participants had little knowledge about DMs. Nevertheless, the teacher managed to consolidate the concept in students' minds given that such features emerged in their speech after a few sessions of treatment. They were also able to use the DMs with more confidence, substituting formal and written DMs such as first, second, etc., with what's more important, next, also and additionally.

Learner's attitude towards the treatment

Follow-up interviews were conducted with 10 participants in the treatment group. Five questions about the treatment were asked and the results were content analysed through thematic analysis as fully described in the method section and the frequency for each theme was calculated.

Question 1 asks, "How did you like this method of teaching speaking?". Five main answers were given to this question, which is summarized in Table 5. Note that the total number of answers exceeded the number of participants as most of them mentioned several points.

Table 5. Emerging themes in the participants' reply to the first question of the interview

Theme	Frequency	Percentage
Increased feedback	8	80%
Improved communication	6	60%

Decreased stress	5	50%
Improved motivation	4	40%
Intimate conversation	4	40%

As presented in Table 4, the participants saw feedback and communication as the two factors which affect their speaking abilities. A majority of the respondents indicated the increased amount of group communication and feedback received by the instructor as the two influencing factors. Some also addressed the intimate conversations among the groups of students as an accelerator for improved communication. They also referred to the decreased amount of stress in comparison to conventional methods of assessment and their motivation to correct their errors for the next assessment as other influential factors.

Question 2 in the interview asks, "What do you dislike about this method?" The answers to this question are reported in Table 6.

The most frequent answer was that the interviewees disliked listening to teachers' comments given to other students. Two interviewees the feedback discouraging and referred to them as finding the faults. Moreover, one interviewee referred to the time consumption of this method in comparison to other methods. A student also referred to the threat of setting the situation for cheating in group assessment.

Table 6. Emerging themes in the participants' reply to the second question of the interview

Theme	Frequency	Percentage
Disliking listening to the teacher's comments Given to other students	4	40%
Time-consumption	3	30%
Discouraging comments	2	20%
Pave the way for cheating	1	10%

Question 3 of the interview asks, "How did you like this as an assessment process?" The categories are reported in Table 7.

Table 7. Emerging themes in the participants' reply to the third question of the interview

Theme	Frequency	Percentage
Helping with learning	7	63.6
More accurate assessment	3	27.3%
Difficulty of application	1	9.1%

A majority of the interviewees characterized this method not only as a system of assessment but also a means of improving learning. They compared this method with traditional methods. They believed that traditional methods of assessment, which they had experienced, were taken at the end of the class or sometimes during the class but they did not provide them with information and awareness about their problems and errors. They also referred to the feedback they received as influential in improving their learning as well as knowing their current situation of knowledge. Moreover, comparing the method with traditional static ones, they were happy that their performances are not measured in one shot but throughout the sessions. This, according to them, gives a more accurate estimation of their abilities. There was one case that referred to the method as "not possible to be used in most educational contexts". It was elaborated by the interviewee that as this method is both time-consuming and requires lots of effort, it cannot be considered as an applicable method to educational contexts, especially the context of Iran which is based on static assessment.

Question 4 of the interviews asks, "What are the advantages or disadvantages of this teaching speaking practice in comparison to other classroom tasks?" The answers to this question were coded and categorized. Table 8 presents the frequency results.

Table 8. Emerging themes in the participants' reply to the fourth question of the interview

Theme	Frequency	Percentage
Feedback	7	35%
Dynamicity of assessment	4	20%
Helping with self-regulation	3	15%
Time management	3	15%
Availability for future references	2	10%
Improving confidence	1	5%

The advantages perceived by the interviewees can be analyzed from two perspectives: a) advantages of dynamic assessment and b) advantages of online classes. With regard to the advantages of dynamic assessment, the most frequent advantage addressed was the feedback received from the teacher. The regular communication and attempts for improvement based on feedback were very helpful in their learning. They also emphasized how the dynamicity of assessment helps them with learning and makes a friendly environment and does not merely judge them as it happens in static assessments. With regard to the advantages of online classes, time management was the most frequent issue addressed. They referred that there was no need for commuting between home and class, and they could access the classes on their phone wherever they wanted. Also, as all materials and comments were available online, accessing them in the future was available. Moreover, one shy student believed that it made me feel more confident because of the online environment.

One issue, both related to the dynamic assessment and online classes, was helping with self-regulation. The dynamicity of the assessment helped the learners to be aware of their problematic areas and the online environment kept the needed material for future learning. However, it was the learners' responsibility to use both and take care of their learning. Finally, question 5 asks, "What are the potential drawbacks and shortcomings of this teaching practice?" The answers are summarized in Table 9.

Table 9. Emerging themes in the participants' reply to the fifth question of the interview

Theme	Frequency	Percentage
Problems with internet/software	8	61.5%
Time consumption	2	15.4%
Preferring face-to-face interaction	3	23.1%

Most of the participants addressed problems such as poor internet connections, running out of battery, and not being able to work with the software as it was new to them. Few learners believed that the process of dynamic assessment is time-consuming and prevents them from learning the materials as fast as they were used to learning in other classes. One learner also preferred to have face-to-face interaction as she was willing to extend her social life. The results of the independent samples t-test indicate a significant difference between the experimental and the control groups with the former doing better in the post-test. Hence, it can be concluded that using MALL-based GDA has been effective in improving the learners' speaking ability. Particularly, the use of cumulative GDA based on the interactionist model has been effective in developing Iranian EFL learners' use of discourse markers.

The findings of the study were consistent with previous findings showing that EFL students who had received GDA as the treatment improved their language performance (Ashraf et al. 2016; Tarighat & Khodabakhsh 2016). The findings of the current study are partially in line

with those obtained by Kazemi and Tavassoli (2020). In addition, the results of the present study were congruent with the conclusions drawn from the study conducted by Safdari and Fathi (2020) revealing that the DA promotes the participant's speaking accuracy and fluency. The studies done by Moradian, Asadi and Azadbakht (2019), and Ebadi and Asakereh (2017) are also among the relevant research reporting the effectiveness of GDA and DA.

According to the results obtained from the interview, which aimed to understand the participants' overall impression of GDA, students' impression was positive about the course which they had passed. Based on the interviewees' opinions, their overall positive attitude was due to the feedback given to them during GDA sessions. As mentioned by Poehner (2005), GDA is intended to check students' possible improvement and support them to achieve this improvement by offering different types of mediations consistent with the concept of ZPD proposed by Vygotsky.

Some students noted that the communication they had with the teacher was greatly useful for their learning. They stated that all of the students had the opportunity to cooperate in the classroom and complete the tasks. Some others also mentioned that GDA is a more accurate type of assessment compared to the traditional type in that it is not one-shot, and it is informative and reveals the problem. These are in concordance with the findings of Poehner (2007) which emphasized that DA isn't an independent activity completed in isolation from other instructive exercises. In other words, based on Poehner (2005), DA is a dialectical method of learning a second language.

The above information, obtained from the interviewees' viewpoints, is also in agreement with Aljaafresh and Lantolf's (1994) assertion emphasizing the gradual and continuant process of providing mediations from the most implicit to the most explicit one which enables the mediator to identify the border of ZPD for each learner in the classroom. When the interviewees were asked about the advantages of online GDA, they hinted at the usefulness of mediation provided by the teacher. Concordantly, Lidz (1991) believed that if mediations cater to each person's needs, each learner's ability improves through this kind of assessment.

A few interviewees said that they did not like listening to the teacher commenting on other students' problems. In contrast, some other interviewees claimed that paying attention to others' problems and the way the teacher helped them to overcome their problems were supportive of mastering what they knew before or gradually solving their problems which were similar to the problems of other students. These findings corroborate the ones achieved by Miri et al. (2017) supporting that "learners need to realign their objectives" and take part in an apparent or covert fashion when sessions with a focus on ZPD are conducted (p. 20).

CONCLUSION

The purpose of the study was to investigate the impact of GDA with a focus on discourse markers on enhancing the speaking abilities of Iranian EFL learners using both quantitative and qualitative methods. It was found that cumulative GDA with discourse markers in focus has a positive impact on Iranian EFL learners' speaking. The results indicate that such a blend of teaching and testing can be useful in educational settings based on distance learning. Previous studies such as Abdulaal et al. (2022), Alavi et al. (2012), Birjandi et al. (2013), Farahani and Moghadam (2018), Rashidi and Bahadori Nejad (2018) Yakışık and Çakır (2017) also confirmed the usefulness of DA, GDA or cumulative GDA. In terms of discourse markers and mobile-assisted language learning in distance learning, the results of this study cannot be compared to previous studies based on these variables given that their interplay was not investigated in previous studies. However, the positive effect of DM teaching on the speaking proficiency of the experimental group participants in this study is in contrast with the results of Khandaghi Khameneh and Fakhraee Faruji (2020) in which they could not establish any relationship between these two variables. This could be because dynamic assessment and

distance learning also positively intervene in the case of DM teaching.

The findings of the current study can be beneficial for teachers, learners, curriculum designers, and pedagogical administrators. Given that discourse markers in spoken English play an important role and bear a good portion of assessment in high-stakes tests such as FEC, IELTS, and TOEFL, the results of this study could inspire textbooks and teacher educators to incorporate teaching DMs based on DA or GDA in educational textbooks as well as pre-and in-service courses. In the teacher-learner interactions, the learners become not only aware of their problems but also are supported to solve their problems and perform beyond their present capabilities (i.e., ZPD). Being exposed to GDA, learners also have the opportunity to overcome their errors while there are interrelations of their goals with the classroom's common goal. The syllabus designers may benefit from the results of the study by designing certain textbooks which focus on GDA. They can also align the course description, purpose, outcomes, and goals with the guidelines of GDA so that they can unify instruction and assessment into a single textbook. By doing so, the teachers will not have concerns about designing their own activities which focus on both instruction and assessment.

While this study provided empirical evidence for the effectiveness of blended teaching and assessment in distance education, it did not investigate the attitudes of teachers implementing dynamic assessment in distance education based on flipped learning and blended education. Comparative studies are also needed to compare the effect in different educational settings including a comparison between educational settings in distance education versus those in traditional face-to-face instruction. It is hoped that this study would be able to open new research paths in EFL/ESL language testing and teaching and encourage research-informed practices with a focus on assessment for learning in distance education. The current study was undertaken among Iranian EFL learners at the B2 level of proficiency; some parallel studies are required to examine the role of cumulative GDA in improving the use of discourse markers among EFL learners at other levels of language proficiency. Further research studies may also probe the effect of the other type of GDA named concurrent GDA, considering interactional and interventional models. Also, it is suggested that further studies could investigate the role of cumulative GDA in enhancing other language skills or sub-skills. Ultimately, language testing has long been considered as a rather separate subfield of TESOL, though research on dynamic assessments including this study, paves the way for debunking the fallacy that assessment and teaching cannot be merged to the benefit of learners. Hopefully, in the near future, testing will be a "celebration of learning" fostering the language abilities of learners through engaging tasks. Such great achievement can only be possible through implementing advanced educational theories and approaches along with fully flourishing the potential of distance education to foster students' proficiency in language and other school or university subjects.

ACKNOWLEDGMENT

References

[1] Abdulaal, M. A. AD., Alenazi, M. H., Tajuddin, A. J. A., & Hamid, B. (2022). Dynamic vs. diagnostic assessment: Impacts on EFL learners' speaking fluency and accuracy, learning anxiety, and cognitive load. *Language Testing in Asia,* 12 (32). https://doi.org/10.1186/s40468-022-00179-0

[2] Aijmer, K. (2002). *English Discourse Particles*. Amsterdam: John Benjamins.

[3] Alavi, S. M., Kaivanpanah, S., & Shabani, K. (2012). Group dynamic assessment: An inventory of mediational strategies for teaching listening. Journal of Teaching Language Skills, 3(4), 27-58.

[4] Alemi, M., Miri, M., & Mozafarnezhad, A. (2019). Investigating the Effects of Online Concurrent Group Dynamic Assessment on Enhancing EFL of Accuracy Grammatical sLearnerInternational Journal of Language Testing, 9(2), 29-43.

[5] Aljaafreh, A., & Lantolf, J.P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The modern language journal, 78*(4), 456-483

[6] Badhoda, I., & Shabani, K. (2018). Response latency as a tool to study L2 learners' ZPD, ZAD, and ongoing information processing. *Asian-Pacific Journal of Second and Foreign Language Education, 1*(2), 1-16.

[7] Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualizing assessment feedback: A key to improving student learning? *Studies in Higher Education, 36*(6), 671–687.

[8] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.

[9] Birjandi, P., Estaji, M., & Deyhim, T. (2013). The impact of dynamic assessment on reading comprehension and metacognitive awareness of reading strategy use in Iranian high school learners. *Iranian Journal of Language Testing, 3*(2), 61-77.

[10] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101.

[11] Creswell, J. W., & Plano Clark V. L. (2007). *Designing and conducting mixed methods research.* SAGE.

[12] Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE.

[13] Ebadi, S., & Asakereh, A. (2017). Developing EFL learners' speaking skills through dynamic assessment: a case of beginner and an advanced learner. Cogent education journal, 4(1).

[14] Farahani, F., & Moghadam, F. S. A. (2018). The Impact of cumulative group dynamic assessment on the learning of congruent and non-congruent collocations among Iranian intermediate EFL learners. *Journal of Applied Linguistics and Language Research*, 7 (4), 21-36.

[15] Farangi, M. R., & Kheradmand Saadi, Z. (2017). Dynamic assessment or schema theory: The case of listening comprehension. *Cogent Education*, *4*(1), 1312078. https://doi.org/10.1080/2331186x.2017.1312078

[16] Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogical settings. Applied Linguistics, 28 (3), 410-439.

[17] Hidri, S. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context. *Language testing in Asia, 4*(1), 1-19

[18] Khandaghi Khameneh A & Fakhraee Faruji L. (2020). The Effect of teaching Discourse Markers (DMs) on speaking achievement among Iranian intermediate EFL learners.

IJREE. 5 (4).

[19] Köroğlu, Z. C. (2019). Interventionist dynamic assessment's effects on speaking skills testing: case of ELT teacher candidates. *Advances in Language and Literacy Studies, 19*(3).

[20] Kumaravadivelu, B. (2005). *Understanding language teaching: From method to post-method*. Mahwah, NJ: Erlbaum.

[21] Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V. S. Ageyev., & S. M. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 99-116). Cambridge: Cambridge University Press

[22] Lidz, C.S. (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford Press.

[23] Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, *9*(3), 3351–33514.

[24] McNamara, T, & Roever, K. (2006). *Language testing: The social dimension*. Oxford: Blackwell.

[25] McNamara, T. (2000). Language testing. Oxford: Oxford University Press.

[26] Miri, M., Alibakhshi, G., Kushki, A., & Salehpour Bavarsad, P. (2017). Going beyond one-to-one mediation in zone of proximal development (ZPD): Concurrent and cumulative group dynamic assessment. *Eurasian Journal of Applied Linguistics, 3*(1), 1–24.

[27] Moradian, M.R, & Baharvand, P. (2015). The effect of group dynamic assessment on raising young Iranian EFL learners' metacognitive awareness and listening comprehension. *English Language Teaching, 2*(3), 67-86.

[28] Nazari, B., & Mansouri, S. (2014). Dynamic assessment versus static assessment: A study of reading comprehension ability in Iranian EFL learners. *Journal of Language and Linguistic Studies*, 10(2), 134-156.

[29] Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*(1). https://doi.org/10.1177/1609406917733847

[30] Ouden, M. D., Keuning, J, & Eggen, T. (2020). Fine-Grained assessment of children's text comprehension skills. In F. Goldhammer, R. Scherer, & Greiff, S. (Eds.). (2020). *Advancements in technology-based assessment: Emerging item formats, test designs, and data sources* (pp. 171, 183). Frontiers Media SA.

[31] Peraza, P. G. (2019). *Guidelines for conducting, transcribing, and analyzing interview*. Retrieved from: https://csass.ucsc.edu/certification/peraza.pdf.

[32] Poehner, M. E. (2005). *Dynamic assessment of oral proficiency among advanced L2 learners of French*. The Pennsylvania State University.

[33] Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly, 43*(3), 471-491.

[34] Rashidi, N., & Bahadori Nejad, Z. (2018). An investigation into the effect of dynamic assessment on the EFL learners' process writing development. *SAGE Open*, *8*(2), 215824401878464. https://doi.org/10.1177/2158244018784643

[35] Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. Routledge. https://doi.org/10.4324/9781315656762

[36] Shrestha, P., & Coffin, C. (2012). Dynamic assessment tutor mediation and academic writing development. *Assessing Writing, 17*,55-70.

[37] So, H. J., Jihyang, L., & E. Lee. (2021). Multi-layered ecological structure of blended learning in science and engineering education in Korea. In Ch. P. Lim, & Ch. R. Graham, (Eds.), *Blended Learning for Inclusive and Quality Higher Education in Asia* (pp. 171-194). Springer.

[38] Son, G., & Kim, S. (2017). *The Potentials of Dynamic Assessment for the Development of English Speaking Performance: A microgenetic analysis*. Retrieved from http://scholar. dkyobobook.co.kr/searchDetail. laf?barcode=4010025100380

[39] Tabatabaee, M., Alidoust, M., & Sarkeshikian, A. H. (2018). The effect of interventionist and cumulative group dynamic assessments on EFL learners' writing accuracy. *Applied Linguistics Research Journal, 2*(1), 1-13.

[40] Tarighat, S., & Khodabakhsh, S. (2016). Mobile-assisted language assessment: Assessing speaking. *Computers in Human Behavior*, 64, 409–413. https://doi.org/10.1016/j.chb.2016.07.014

[41] Yakışıki, B. Y, & Çakır, A. (2017). Dynamic assessment of prospective English teachers' speaking skills. *European Journal of English Language Teaching*, 2(1), 22–53.