The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety

Nasiba Sherkuziyeva[1] , Farida Imamutdinovna Gabidullina[2] , Khaled Ahmed Abdel-Al Ibrahim[3,4] and Sania Bayat[5*]

*Correspondence:
Sanybayat5656@gmail.com;
saniabayat@yahoo.com

[1] The Department of Corporate Finance and Securities, Tashkent Institute of Finance, Tashkent, Uzbekistan
[2] Department of Tatar Philology, Candidate of Philological Sciences, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia
[3] Educational Psychology, College of Education, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
[4] Sohag University, Sohag, Egypt
[5] Department of English Literature, Faculty of Foreign Languages, Rasht Branch, Islamic Azad University, Rasht, Iran

Abstract

This study aimed to examine the impacts of computerized dynamic assessment (C-DA) and rater-mediated assessment on the test anxiety, writing performance, and oral proficiency of Iranian EFL learners. Based on Preliminary English Test (PET) results, a sample of 64 intermediate participants was chosen from 93 students. Running a convenience sampling technique, target test-takers were randomly divided into the experimental groups (C-DA) and control (rater mediated assessment). Following that, both groups had pretests for oral and written skills. The Science Anxiety Scale (SAS) was also used to gauge their level of anxiety prior to treatment. The experimental group's participants then received C-DA. Rater-mediated assessment, on the other hand, was given to the control group. Both groups took the post-test for writing performance, oral proficiency, and test anxiety at the conclusion of the treatment. According to the one-way ANCOVA analysis, the post-test results for the two groups were different. Based on the results, the experimental group outdid the control group on the oral proficiency, writing performance, and test anxiety post-tests. Iranian EFL learners were able to improve both their written and oral skills while experiencing less test anxiety thanks to C-DA. Finally, the conclusions, the implications, the limitations, and the suggestions for further studies were provided.

Keywords: Computerized dynamic assessment, EFL learners, Oral proficiency, Rater mediated assessment, Test anxiety, Writing performance

Introduction

Assessment is a planned process in which educators use data on learners' progress to modify their continuing teaching methods or learners use it to modify their current instructional approaches (Popham, 2008). In the field of learning and teaching, assessment is a technique used by teachers and pupils during teaching to offer the required feedback to change ongoing learning and instruction to advance learners' achievement of intended goals (Robinowitz, 2010). Assessment seeks to enhance

learning and bridges the gap amid learners' current learning environment and their desired learning purposes (Heritage, 2012).

C-DA and rater-mediated assessment are two forms of assessment. The socio-cultural theory (SCT)-based dynamic assessment (DA), an alternate to traditional assessment, emphasizes the process-oriented elements of learning (Carney & Cioffi, 1990). Lev Vygotsky created the SCT, a psychological theory whose principles form the cornerstone of DA's reliance on the union of evaluation and teaching. If we include a mediation phase in our assessment according to DA, development will take place (Lidz & Gindis, 2003).

Regarding this, Vygotsky (1978) described mediation as the appropriate form of assistance that fosters students' skills and asserted that it would be more effective if it is centered on each learner's Zone of Proximal Development (ZPD), that is the distance between an individual actual and potential capacity. C-DA, a continuous component of DA, offers learners computer-assisted automated mediations. The use of C-DA has some benefits, including the ability to assess many students at once, allowing students to retake the test as often as they like, and making a scoring file for every student as they complete the required tasks (Ebadi & Saeedian, 2015).

Rater-mediated assessment is another kind of assessment. Assessments that are "rater-mediated" are those by which the raters assess performances of participants and utilize classifications on a rating scale to represent the degree of performances on one or more areas. When describing the levels of test-taker performances using rating-scale categories, raters frequently refer to rubrics and performance-level descriptors for guidance. Human judgements nevertheless direct the creation and assessment of the algorithms that eventually score test-taker responses, even when automated scoring processes are employed (Engelhard Jr & Wind, 2019).

To evaluate test-taker performances in a number of contexts and subject areas, including educational performance assessments, language proficiency tests, and personnel evaluation, researchers and practitioners from all over the world employ rater-mediated assessments. In general, individuals choose rater-mediated tests because they think they give more relevant information into test-taker locations on a certain concept than examinations that can be scored without rater assessments (Engelhard & Wind, 2017).

Among all, writing is a dynamic ability that is now valued more and more in all facets of education and communication. Writing is always crucial in the study of foreign and second languages. An individual who is skilled in listening, reading or speaking in a foreign language is not considered to be an adept language learner nowadays unless she or he also possesses reasonable writing abilities in the specific language. It is a potent communication method that fosters critical thinking and makes learning easier (Biancarosa & Nair, 2007).

Additionally, a language learner's academic success and growth in all subject areas frequently depend on their capacity to communicate their knowledge in writing (Valasa et al., 2009). Thus, writing has become increasingly significant in the study of and instruction in second and foreign languages (Khodashenas & Rakhshi, 2017). Furthermore, regarding oral skills, many approaches to conceptualizing oral proficiency have been used in studies on the acquisition of the English language. It entails receptive and expressive abilities, as well as understanding or usage of certain parts of spoken

language, that pertains to syntax, vocabulary, phonology, morphology, pragmatic skills, and discourse characteristics (August & Shanahan, 2006).

The ability to communicate orally is critical to language students since it ultimately is the skill that is mostly utilized. The majority of our daily exchanges are verbal. Therefore, it is crucial for second and foreign language learners to strengthen their spoken language skills. Learning to speak English fluently requires not only expanding one's vocabulary, mastering grammar, and comprehending the language's complex and subtle semantics, but also learning how to communicate effectively with native English speakers (Genesee et al., 2006).

Another variable, test anxiety, relates to how much fear, worry, uneasiness, panic, restlessness, and tension students feel when just thinking about upcoming examinations or tests. Another way to think of anxiety is as a result of uncertainty about impending things (Craig, 1995). Test anxiety is the term for the emotional reactions or feeling of uneasiness that develops prior to exams and lasts during the exam period (Sepehrian, 2013). Anxiety is frequently linked to self-efficacy pressures, assessments of the severity of the situation, and responses to a source of stress (Pappamihiel, 2002).

Review of literature
Theoretical background
Although DA has carved out a significant place for itself in developmental psychology, L2 researchers have only just begun to take notice (Mallahi & Saadat, 2020), in part due to its sound theoretical foundation and potency in helping students acquire new cognitive skills. Due to a number of factors, including a focus on process not product, the incorporation of assessor feedback, and a change from examiner impartiality to a tailored instruction and supporting relationships, DA replaced the typical static assessment (Sternberg & Grigorenko, 2002). DA differs from conventional assessment methods since it is predicated on the idea that human skills are dynamic and unstable, as opposed to static. DA openly criticized IQ-based testing for their propensity to just provide a fixed estimate of students' capabilities (Pishghadam et al., 2011).

DA is an integrated method of assessment and instruction that is unexpected, constantly changing, and dynamic and is derived from Vygotsky's sociocultural theory (SCT). DA is a strategy to comprehend individual differences and their consequences for instruction that embeds intervention inside the assessment system, acknowledging the interrelationship between learning and assessment (Zangoei et al., 2019).

DA is a process that aims to simultaneously assess and advance learners' capabilities. Additionally, it promotes learners' independence in problem-solving and knowledge construction and, as a substitute to traditional assessment procedures, it incorporates mediation into the assessment process as an instrument to determine their full capacities and foster their emerging skills (Poehner & Lantolf, 2005).

DA typically consists of three phases: test, teach, and re-test. In the test phase, the examiner observes the testee's individual abilities in a task with little to no help from the testee; in the teach phase, the examiner helps the testee in tasks that are parallel to the ones used in the test phase; and in the re-test phase, the testee is tested independently once more. Changes made between test and retest phases serve as a measure for the mediation's effectiveness. Therefore, interventionist and inter-actionist are the

two main methods of DA (Davoudi & Ataie-Tabar, 2015). Lantolf and Poehner (2004), in developing a theoretical basis for DA processes, labeled different forms of mediation as interactionist or clinical DA and intervention or psychometric DA, respectively. The key distinction between the two models is how the mediation is delivered to pupils. One distinguishing element is that mediation between the student and the instructor can be negotiated (interactionist) or predetermined (interventionist).

C-DA have lately acquired traction in the context of second/foreign language (L2/FL). C-DA, which is based on Vygotskian socio-cultural theory, combines instruction and assessment by offering adapted electronic mediations to pupils. This alternative viewpoint is founded on Vygotsky's theory of socio-cultural and his learner-centered concept of DA, in which the instructor engages or meddles in the pupil activity in order to broaden and expand her/his learning potential while also diagnoses or evaluates their development (Van der Veen et al., 2016). This model aims to examine diverse skills, and the erroneous response provides an instructional program that re-evaluates the given subjects using computer technologies. In essence, C-DA emerged from computerized testing, which sought to make up for the shortcomings of old paper-and-pencil testing and to improve L2 learners' collaboration with examiners by fostering a non-threatening and learning-focused environment (Poehner & Lantolf, 2013).

The Vygotsky's ZPD is the model's primary underlying principle. To determine an individual's potential degree of development, ZPD employs diagnostic concepts. The SCT epistemology of Vygotsky is reflected in the ZPD theory, which holds that teaching and evaluation can be linked so as to provide a comprehensive picture of the pupil's underlying potential (Shabani, 2021). According to Vygotsky (1978), all learning occurs in the "zone of proximal development," that symbolizes the space between mediated and unmediated capacity (Hasson & Botting, 2010). The idea behind ZPD is that by providing proper scaffoldings and mediation, learners' word recognition capacity may be maximized (Vygotsky, 1978).

ZPD is a cornerstone of sociocultural theory (SCT), which is used as a diagnostic instrument for both educators and researchers to get a clearer understanding of how learners grow and the kinds of challenges that impede cognitive development. The difference among the real developmental level as measured by self-directed problem solving and the level of potential development as shown by problem solving under other supervision or in collaboration with more proficient peer is described as ZPD (Vygotsky, 1986).

Multimedia and computer programs are two instances of artifact mediation that have the potential to be very effective at scaffolding L2 learning (Laufer & Hill, 2000). The SCT-based methodological process known as DA captures ZPD the best. DA distinguishes between unmediated and mediated (ZPD) performance of learners. C-DA includes help and assessment to resolve the pupils' language deficits (Shabani, 2021).

In rater-mediated assessments, on the other hand, the testing setting is defined by interactions between the rater, the rating scale, rating procedures, and learners' achievement (McNamara, 1996). Rater variability, defined as the way(s) raters may generate construct-irrelevant variance in the points granted to students' second language performance, is a main consideration in this assessment (Myford & Wolfe, 2003). Learners' written or spoken scores may not necessarily represent their language competence since

a variety of other construct-irrelevant elements may impact the knowledge they exhibit. Rating scales and rater subjectivity are two variables that might have an impact on the final results (Esfandiari, 2021).

The rating process in rater-mediated assessments is a difficult undertaking in which rater subjectivity may contribute to construct-irrelevant variation (Wind, 2020). According to Cronbach (1990), when raters engage in the rating process, they are involved in an error-prone and complicated cognitive process in which they may not logically follow the rating criteria confidently and constantly, even after extensive rater training (Knoch et al., 2020).

Rating scales, in addition to raters, have an impact on ratings. Myford and Wolfe (2004) succinctly defined a rating scale as a measurement device used to record the findings of the rater's observations. Eckes (2015), also, observed that the kind of rating scale used affects how language instructors make significant assessment judgments. This judgment is vital because the score is ultimately what will be utilized in making judgment and inferences about learners (Weigle, 2002).

Writing ability, as the first notion to explore, is a producing skill that is influenced by assessments. Writing skills are important in communication because they enable individuals to express their thoughts, feelings, and views. Writing is both a physical and a mental activity. Writing is the act of physically transferring ideas or words to surface. Writing, in contrast, is the mental process of thinking up concepts, deciding on transmitting them, and changing them into words and sentences that a reader can comprehend (Nunan, 2003).

Writing can also be defined as a process, or the processes a writer employs to produce something in its completed form (Harmer, 2006). The four essential components of this process include planning, drafting, editing, and creating the final paper. In this line, writing is perceived as the final result of extensive planning, composing, reviewing, and revising operations (Richard & Schmidt, 2002).

Scholars have recently broadened their perspectives on writing to incorporate a social component. Writing is both a complex social action and a cognitive activity. It demonstrates the author's communication skills and topic understanding. Writing is tough to learn and master, especially in a second language such as English (Shokrpour & Fallazadeh, 2007).

Writing is always the last skill to be learnt, after listening, speaking, and reading. Writing, on the other hand, is viewed as the hardest expertise for learners to acquire. English as a foreign language is no exception. Learners frequently struggle to write texts in their native language. These issues appear to be substantially worse in English writing. This has been a hotly debated matter among foreign language specialists and linguists all over the world (Ngoc Anh, 2019).

Test performance is ascribed to test-task test-taker features. Linguistic knowledge, topical knowledge, personal qualities, emotional schemata, and strategic competency are all test-taker characteristics. The first three interact with the last two, and test performance is the result (Bachman & Palmer, 1996).

As a result, test-taker and test-task qualities influence one another. Understanding how these components impact test performance is imperative since teachers' judgments based on test results rely on these qualities. While all of these variables merit

examination, one fundamental question is how personal qualities influence test performance. Test anxiety is the most important of all personal factors associated to test performance (Bachman & Palmer, 1996).

According to Sarason (1988), anxiety is a basic human feeling that is comprised of dread and insecurity and manifests itself when anything appears to be a danger to the ego or self-esteem. Test anxiety has also been introduced as a particular type of general anxiety characterized by physiological, phenomenological, and behavioral reactions associated with the fear to fail and the experience of testing or assessment (Sieber, 1980).

Spielberger (1972) was the first academic to distinguish between two forms of anxiety: trait anxiety and state anxiety. The latter is, regrettably, considerably more common among teenagers, particularly in academic settings. State anxiety is a temporary feeling of unease accompanied by physical and behavioral responses triggered by the autonomic nervous system. Test anxiety is a type of state anxiety since it arises only when people are being tested and their performance is crucial. Wren and Benson (2004) classified test anxiety into three components: thoughts like self-criticism and performance concerns, autonomic responses like sweaty hands, elevated heart rate, dry mouth, headache, and off-task behaviors like fidgeting.

Similarly, speaking abilities are essential components of learning a language. Yet, speaking abilities assessing is difficult since there are several elements impacting the assessment procedure (Luoma, 2004). Several areas of speaking ability need to be evaluated, including syntax, lexicon, pronunciation, fluency, and accuracy of expression (Madsen, 1983).

Fluent, competent, proficient, and bilingual are all terms used to describe proficient foreign language speakers. Foreign language competency is influenced by a number of factors. In order to provide a comprehensive vision of the complicated relationships amongst features, four essential traits of fluency, syntactic complexity, lexical variety, and grammatical correctness are crucial (Iwashita, 2010).

The range of forms that appear in language production and their level of sophistication are referred to as syntactic complexity. The length of the production unit, the level of embedding, the degree of coordination and subordination, the diversity of structural types, and the structural intricacy are all quantified by researchers (Iwashita, 2010).

Lexical richness is referred to as lexical variety. Grammatical accuracy is related to both general accuracy (the recognition of all forms of mistakes) and particular categories of flaws. Fluency refer to the temporal characteristics of speech (syllables and words per minute, pauses number or length) and the automaticity with which language is used (how well pupils can produce a L2 without attending to grammar rules) (Iwashita, 2010).

Empirical background

C-DA implementations for assessment and educational objectives have lately been documented in L2 teaching scholars various works. For instance, Hidri and Roud (2020) looked at the effects of C-DA on a TOEFL iBT reading exam among 185 Iranian students who were upper-intermediate EFL speakers. Findings showed that utilizing hints in the question types resulted in statistically significant changes between real and mediated scores with different degrees of reading ability. Significant correlations were produced by C-DA, which raised the test results on the mediated items.

In a quasi-experimental design, Yang and Qian (2019) employed C-DA as assessment and teaching approach to improve the reading comprehension of Chinese EFL learners. Both the control and experimental groups took three exams, although in different forms. The findings showed that after 4 weeks of instruction, the experimental group scored much better than the control group, despite the fact that reading comprehension ability in the two groups was very equal at the beginning of the research.

Moreover, Bakhoda and Shabani (2016) investigated the learners' processing time and response latency (RL) throughout C-DA in reading comprehension to increase C-DA application to the field of cognitive psychology. Based on the processing time needed for each mediation, the devised program could identify between students with higher and lower ZPDs. Zhang and Lu (2019) argued that the diagnostic data from C-DA might be applied to the classroom to direct instructional activities.

Poehner et al. (2015) broadened C-DA application to reading and listening abilities and presented various results for all students, containing real scores, transfer scores, mediated scores, and learning capacity scores. They claimed, however, that their intended mediations inventory is not effective for students in different settings because the work might be difficult for them.

The study by Pishghadam and Barabadi (2012) supported the construct validity of C-DA in revealing students' learning capacity, which was underdeveloped in their initial unmediated capability. The effects of C-DA on cognitive performance in comparison to DA with an examiner were investigated by Tzuriel and Shamir (2002). The results showed that mediation processes in a C-DA procedure were more successful in producing significant cognitive changes than mediation involving just an examiner.

Several studies also show consistent correlations among text sophistication and rater assessment of writing accomplishments, with higher degrees of linguistic sophistication and word numbers positively related to essay ratings. Yang et al. (2015) investigated the link between writing quality as rated by human raters and ESL writing's syntactic complexity, as well as the importance of theme in the relationship. The topic was discovered to have a substantial influence on the syntactic complexity characteristics of the essays, with one theme evoking more subordination (finite and non-finite) and more general sentence complexity and the other generating more explanation at the finite phrase levels (in particular, complex noun phrases, and coordinate phrases). Local-level complexity traits that were more prevalent in essays on a single topic (for example, subordination and elaboration at the finite phrase level) did not correspond with topic scores. Rather, a reversal was observed: the less prevalent local-level complexity characteristics for essays on a single topic tended to have a stronger connection with that topic's ratings.

Kobrin et al. (2011) demonstrated in their study that essays length is connected to scores, however the association is not as strong as earlier critics believed. The association between essays length and performances was dramatically varied between prompts, which was clarified by examinees' average SAT Critical Reading performance for the question.

Similarly, other scholars have investigated the link between essay length and assessed writing achievement, with longer essays often receiving higher ratings. Chodorow and Burstein (2004) investigated the relationship between essay length and holistic scores awarded to TOEFL essays using e-rater, an automated scoring system

established by Educational Testing Service (ETS). The results demonstrated that an early version of the system, e-rater99, explained minimal variation in human reader ratings beyond what might be anticipated from essay length.

A subsequent version of the system, e-rater01, outperformed its predecessor and was less reliant on length due to a greater focus on measurements of topical information, complexity, and variety of words. Essay length was also looked into as a potential reason for disparities in results among examinees who spoke Spanish, Arabic, or Japanese as their first language. Even when the effects of length were controlled, human readers and e-rater01 exhibit the same pattern of disparities for these groups (Chodorow & Burstein, 2004).

Several scholars have investigated the link among textual features of students' essays and the ratings given to them in the setting of rater-mediated assessments for writing. Overall structure, utilization of supporting resources, significant insights, rhetorical technique, and thesis statement were shown to be the strongest predictors in Breland et al. (1995) study.

Significance and objectives of study

The objective of combining diverse assessment methods, such as DA and rater mediated assessment, is to concurrently measure and alter cognitive functioning through mediation and intervention and to move the emphasis to more learner-centered assessments. The end result of DA is an estimation of learning potential rather than a judgment of how smart the examinee is in comparison to their classmates. The premise underlying the learning potential construct is that students with comparable initial capabilities (and hence comparable IQs) may react to instruction differently.

It has been noticed that traditional methods of assessment are used in EFL classrooms in Iran, with the instructor presenting topics to the learners and later evaluating the papers and returning them to the learners, who might not even look at the papers, let alone amend their problems. Old-style instruction and assessment methods in Iran scarcely fit the demands of today's pupils. The fundamental cause of learners' discontent can go back to the distinctive character of conventional evaluation methods. In contrast, the novelty of alternative ways of assessment made it unavoidable for instructors and researchers in Iran to seek to take advantage of such choices.

In conclusion, the preceding and several other researches verified the favorable impacts of novel types of assessment on language acquisition. However, there have been no studies to simultaneously evaluate the impacts of C-DA and rater-mediated assessment on Iranian EFL writing performance, test anxiety, and oral proficiency. To close the current gap, the following research questions were raised:

1. Does applying rater mediated and C-DA affect EFL learners' writing performance differently?
2. Does applying rater mediated and C-DA affect EFL learners' test anxiety differently?
3. Does applying rater mediated and C-DA affect EFL learners' oral performance differently?

Based on these research questions, the following null hypotheses were formulated in this research:

> H01: C-DA and rater mediated assessment do not have any substantial effect on Iranian EFL learners' writing performance.
>
> H02: C-DA and rater mediated assessment do not have any substantial effect on Iranian EFL learners' test anxiety.
>
> H03: C-DA and rater mediated assessment do not have any substantial effect on Iranian EFL learners' oral proficiency.

Methodology

Participants

According to the findings of Preliminary English Test (PET), 64 students were chosen for the research from a pool of 93 Iranian EFL students. They were picked from a private English Language Institute in Ahvaz, Iran. They were 16–23 years old male students with intermediate level. Using a convenience sample approach, we randomly selected respondents and distributed them into two groups: Experimental group (EG): (C-DA) and Control group (CG): (rater mediated assessment). Because of the institute's gender segregation, we could only select male participants.

Instruments

Preliminary English Test (PET)

To begin, Preliminary English Test (PET) was adopted and administered to the individuals during the first session to establish language proficiency homogeneity. PET is a language proficiency test prepared for individuals that can use spoken and written English in intermediate level. This test includes four sections of listening, speaking, reading, and writing. All sections were administered to homogenize the participants. The test was administered to 93 pupils, and after analyzing the results, those individuals with extreme scores were eliminated from the research. The researcher was then certain that all of the study participants had intermediate English language competency at the start of the investigation.

Science Anxiety Scale (SAS)

The participants' test anxiety was measured using the Science Anxiety Scale (SAS) developed by Britner and Pajares in 2006. Some wordings of the items were altered to make them more acceptable in assessing test anxiety. This 12-items test requested participants to consider the items (like I am worried I will receive low scores on most of the examinations) and to respond on a 5-point scale extending from absolutely false to absolutely true. SAS has a reliability index of .79 according to the Cronbach's alpha formula. Its validity was confirmed by three experts in English teaching.

Writing Scale

A writing pre-test developed by researchers was used to collect the data needed to reply to the writing questions. The pretest was according to the course book (practical writer

with readings). Students were required to choose one of two subjects and write an essay about it. The learners were required to prepare a composition on the chosen subjects by the direction of the researcher. Participants were also required to include at least 100 words in their writings. The researcher observed the administration of the pre-test in the classroom to confirm that the students completed it on their own. Following the completion of the writing assignments, all essays were collected and scored using the academic IELTS writing criteria. Pearson correlation analysis was used to determine test reliability (= 0.89). In addition, the pre-test validity was certified by two specialists in English.

The current investigation also included a post-test in writing. On the post-test, learners were asked to produce a 100-word essay. The student essays were graded by raters. A post-test was administered to test-takers to see how well their writing abilities had enhanced as the consequence of the treatment. It should be stated that the reliability of writing post-test (0.87) was assessed using Pearson correlation analysis. Two experts in the field of English teaching measured its validity. The Academic IELTS writing evaluation criteria was applied to evaluate the participants' writing assignments. Examiners assessed the writing samples using specific performance descriptors based on the rubric. These four performance categories—grammatical accuracy and range, coherence and cohesion, task response, and lexical resource, —were considered by the descriptors.

Oral Proficiency Scale

The other instrument used in this survey was a researcher-made pre-test of oral proficiency that contained numerous items from the learners' course books (i.e., Top Notch 3). The participants got 2 to 3 min to address the subjects. Two raters examined the respondents' speaking performances, while their productions were being recorded for the next rater. Hughes speaking checklist (2003) was used to score the participants' oral productions. The test validity was confirmed by two experts in English teaching. Additionally, the reliability of speaking test was determined to be (= .83) using Pearson correlation analysis. It should be stressed that this test served both as speaking pre-test and post-test.

Procedure

To decide the test-takers' homogeneity in their degree of English proficiency, the researcher administered the PET. Out of 93 participants, 64 were selected to represent the sample subjects in the present study. After that, two equal groups were randomly selected from them (control and experimental Then, both groups took pretests to measure their oral and written competence as well as their test anxiety. After then, various therapies were applied to two groups. C-DA was administered to learners in the experimental group as treatment, whereas rater-mediated assessment was given to individuals in the control group.

The learners were initially introduced to C-DA and rater mediated assessment before using them in both experimental and control groups. A practical software was created to mediate a complex of pre-formulated helpful hints throughout the test administration because one of the study's objectives was to improve the learners' writing abilities.

As this study took the interventionist approach of DA with pretest, instruction (computerized intervention), and post-test procedures, the learners were required to type

their initial draft (independent performance) on computer in one session of 50 min followed by a 15-min break before starting computerized dynamic test of writing (CDTW) in order to manage the time and to lessen the load of typing on the computer. Participants must first type their names or IDs on the software (learner number). Following that, short instructions on how to begin the test are provided to test-takers. Students got the chance to learn from the test throughout this dynamic test. After receiving guidance on selecting, students should follow procedures to complete the work, with recommendations included in each stage.

In the pre-writing phase, to help participants build a network of usable knowledge, information about topics is presented to them through leading questions and infographics. The test takers are given tips to check and modify their answers against a set of standard answers and indications accessible in CDTW after responding and storing the items in order to help them identify essential concepts in arranging their writing. These processes, in all CDTW steps, allow them to think through the process. Meanwhile, while the test-takers are completing the tasks. They are also performing self-evaluation.

In the second phase, namely writing and drafting, three parts of the introduction, body and conclusion of writing were included. For each section, the test-takers had to revise and then save their own early typed draft as their first dependent effort following tip 2 in the introduction, tip 3 in the body, and tip 4 in the conclusion section to improve their writing in organization, reasonable development and content, coherence, cohesion, and quality or style of expression. In this stage, the software develops (offering the participants tips to develop) fluency in respect of coherence, and lexical complexity regarding vocabulary diversity across paragraphs. Students can improve the quality and complexity of their writing by integrating concepts into a complex web of relationships expressed in key phrases or by consulting a dictionary.

At the end of test, in the formulation phase, the software provided learners with a similar model essay written by native or native-like competent writers to help them notice any unique features (form, lexical, content, and discourse) of a typical model essay. These There were features highlighted. Furthermore, they could compare their own work with this improved model using the evaluation criteria. Finally, as their final performance, the participants were required to manually edit their previously stored writing.

The scores assigned to the final modified-writing in handwriting format were accounted for as a marker of the test-takers' progress in implementing the CDTW's tips in each phase. After learners develop problem-solving skills through continuous self-evaluation and self-modification in CDTW, their aptitude for solving problems can be tested in other tasks of a similar kind. The students' writings were graded using the Bailey and Brown (1984) essay scoring criteria.

In reference to the recommendations made by Lantolf and Poehner (2011), the students obtained a C-DA for oral proficiency. The use of equal-sized groups of students, mediation, cooperation, and interaction among pupils and occasionally between instructor and students, providing scaffolding and necessary support in pupils' ZPDs, and cooperation to generate dialogs, incorporating apologies, request, greeting, and refusals were practically all instructional strategies used in the classroom.

The language institute's best-equipped classroom, which included computers for every student, served as the setting. Naturally, some students utilized their own laptops. With

certain offline and online Software like Wufun, Lingua, Rosetta Stone, and others, the tutor created a few virtual dialogs to evaluate learners' improvement in using proper vocabulary and syntax in their real-world talks.

When necessary, the teacher made sure that the pupils made up for their lack of digital literacy, and occasionally, extra instruction were organized to help some of the students become comfortable with utilizing different computer programs. The challenge in this classroom was the instructor's enormous effort to create the educational materials, which included the researched speech acts, and to solve the technical issues both before and during the sessions.

Meanwhile, in the rater-mediated assessment group, specific procedures were used to achieve the study's goals. Prior to the rating process, the researchers were able to hold a training session for raters. The researchers themselves served as the rater trainer. They provided descriptions on how to rate the essays holistically and analytically. Using analytical and holistic rating tools, they evaluated student writings. The holistic rubrics are scales that give general pictures of the writing performance at each stage. The rubric that instructors use to evaluate their students' work on written assignments is modeled around this one. As a matter of fact, it was created by the researchers after conversations with the instructor about the criteria they employ to evaluate written samples. As a result, the recommended rubric had two performance standards: topic comprehension and linguistic accuracy.

The analytical rubric, however, is a modified form of Bachman and Palmer (1996). Their criterion-referenced scale to rate the writing ability now includes a fifth subdomain that was added by the researchers. Context-specific factors served as the motivation for this addition. The final outcome was a 5-point scale that breaks down writing skill into five different categories: content, cohesiveness, syntactic structures, vocabulary, and writing mechanics. There are a number of precise criteria of performance points inside each area that each rater fully comprehends.

The raters were also shown some previously graded material that had been evaluated holistically and critically. They were individually requested to rate several writings both holistically and analytically in order to improve the training program's performance. When raters gave entirely different evaluations, they were asked to explain these wildly unexpected results.

Punctuation, the need for indentation, expressiveness, and elements of a well-written essay including structure, content, transition, and coherence were all taught to pupils in writing classes. Along with these development patterns, the students learned how to write cause-and-effect essays, comparison and contrast essays, and enumeration essays.

After students completed their writings, they were graded. The rating scale's key criteria were as follows: content, mechanics, organization, coherence, cohesion, grammar, and vocabulary. The scale categories were 1 (extremely poor), 2 (poor), 3 (good), 4 (very good), and 5 (excellent). The raters were also required to provide comments on various parts and characteristics of their writings and, if necessary, fix the students' faults. After the data analyses were finished, the raters got comments on their ratings.

In the speaking part, participants' oral responses were recorded, and raters rated each speech sample using a complete list of performance criteria. Precision of speech, fluency, intonation and prosody, lexical and grammatical accuracy, scope of lexical and

Table 1 Writing accuracy descriptive statistics

Dependent variable: writing accuracy post-test			
Group	Mean	Std. Deviation	*N*
Control	11.78	1.77	32
Experimental	15.25	1.72	32
Total	13.51	2.46	64

Table 2 Writing accuracy inferential statistics

Dependent variable: accuracy post					
Source	Type III sum of squares	df	Mean square	*F*	Sig.
Corrected model	255.186[a]	2	127.59	61.38	.00
Intercept	42.96	1	42.96	20.66	.00
Pre	62.67	1	62.67	30.15	.00
Group	160.69	1	160.69	77.30	.00
Error	126.79	61	2.079		
Total	12073.00	64			
Corrected total	381.98	63			

[a] R squared = .668 (adjusted R squared = .657)

grammatical knowledge, and degree of coherence and structure were among the factors considered. The scale scopes were 1 (quite poor), 2 (poor), 3 (good), 4 (very good), and 5 (excellent). Two groups were given a speaking post-test after the treatments, and their oral abilities were evaluated using the previously mentioned grading standards.

After 19 sessions, with two sessions held in every week, writing performance, oral proficiency, and test anxiety post-tests were run. The acquired data were examined using SPSS software, version 22. ANCOVA was utilized to assess the effects of the cited assessments on the learners' test anxiety, oral proficiency, and writing performance.

Results

Both descriptive and inferential data pertaining to the writing performance, test anxiety, and oral proficiency are presented in the results division. The results and data are detailed in the sections that follow:

Table 1 illustrates that the control group's mean score is 11.78 and the experimental group's mean score is 15.25. It seems that the experimental class got higher scores than the control class on the accuracy post-test. A One-way ANCOVA test was administered in the subsequent table to see whether the differences between the accuracy post-tests of two groups were remarkable or not.

The data in Table 2 indicate that Sig is.00, which is less than 0.05; as a result, there were significant differences between the accuracy post-tests of the two groups. On the accuracy post-test, the experimental participants actually outperformed the control participants. The test-takers in the experimental group were able to improve their writing accuracy thanks to the C-DA technique.

Table 3 shows the control group and the experimental group mean scores are 11.62 and 15.31, respectively. It depicts that the experimental participants outflanked the

Table 3 Writing fluency descriptive statistics

Dependent variable: fluency post-test			
Group	Mean	Std. deviation	*N*
Control	11.62	1.33	32
Experimental	15.31	1.85	32
Total	13.46	2.45	64

control participants on the writing fluency post-tests. A one-way ANCOVA test indicates if the differences between the writing fluency post-tests of both groups were substantial or not.

As Sig (.00) is less than 0.05, it is inferred from Table 4 that there are differences between the two groups on the fluency post-tests. On the fluency post-test, the experimental participants actually outperformed the control participants. The application of C-DA assisted EFL students to improve their writing fluency.

According to Table 5, the experimental group's mean score is 15.98, whereas the control group's mean score is 12.09. On the writing complexity post-test, it appears that the experimental participants outperformed the controls. A one-way ANCOVA test was conducted in the following table to see whether the differences in the writing complexity post-tests of the two groups were statistically momentous:

The differences in the complexity post-tests of both the control and experimental groups were significant, as is shown in Table 6, where Sig is.00, that is less than 0.05. On the post-test writing complexity, the experimental students did indeed outperform the control students. The advantages of C-DA technique may be credited for the experimental group's improved performance on the complexity post-test.

Table 4 Writing fluency inferential statistics

Dependent variable: post					
Source	Type III sum of squares	df	Mean square	*F*	Sig.
Corrected model	250.50[a]	2	125.25	59.02	.00
Intercept	50.99	1	50.99	24.03	.00
Pre	32.93	1	32.93	15.52	.00
Group	173.29	1	173.29	81.66	.00
Error	129.43	61	2.12		
Total	11990.00	64			
Corrected total	379.93	63			

[a] *R* squared $= .659$ (adjusted *R* squared $= .648$)

Table 5 Writing complexity descriptive statistics

Dependent variable: complexity post-test			
Group	Mean	Std. deviation	*N*
Control	12.09	1.95	32
Experimental	15.96	1.89	32
Total	14.03	2.73	64

Table 6 Writing complexity inferential statistics

Dependent variable: complexity post-test

Source	Type III sum of squares	df	Mean square	F	Sig.
Corrected model	240.53[a]	2	120.26	31.98	.00
Intercept	453.94	1	453.94	120.70	.00
Pre	.288	1	.28	.07	.78
Group	240.22	1	240.22	63.87	.00
Error	229.39	61	3.76		
Total	13070.00	64			
Corrected total	469.93	63			

[a] R squared = .512 (adjusted R squared = .496)

The mean score for the experimental group is 44.03, whereas the mean score for the control group is 32.28, as shown in Table 7. The experimental group seems to have fared better on the post-test for test anxiety than the control group. The following table was subjected to a one-way ANCOVA test to see if the differences in test anxiety post-tests between the two groups were statistically significant:

As can be seen in Table 8, where Sig is.00, less than 0.05, there were significant differences between the experimental and control groups' test anxiety post-test results. The experimental group students really performed better than the control group students on the post-test of test anxiety. It is reasonable to attribute the experimental group's increased performance on the post-test for test anxiety to the benefits of the C-DA usage.

Table 9 displays that the experimental group's mean score was 16.00 whereas the control group's was 11.59. In comparison to the control group, the experimental

Table 7 Test anxiety descriptive statistics

Dependent variable: test anxiety post-test

Group	Mean	Std. deviation	N
Control	32.28	3.87	32
Experimental	44.03	5.80	32
Total	38.15	7.68	64

Table 8 Test anxiety inferential statistics

Dependent variable: test anxiety post-test

Source	Type III sum of squares	df	Mean square	F	Sig.
Corrected Model	2279.30[a]	2	1139.65	48.37	.00
Intercept	1904.39	1	1904.39	80.83	.00
Pre	70.30	1	70.30	2.98	.089
Group	2143.03	1	2143.03	90.96	.00
Error	1437.13	61	23.56		
Total	96894.00	64			
Corrected total	3716.43	63			

[a] R squared = .613 (adjusted R squared = .601)

Table 9 Oral proficiency descriptive statistics

Dependent variable: oral proficiency post-test			
Group	Mean	Std. deviation	*N*
Control	11.59	1.49	32
Experimental	16.00	1.34	32
Total	13.79	2.63	64

Table 10 Oral proficiency inferential statistics

Dependent variable: oral proficiency post-test					
Source	Type III sum of squares	df	Mean square	*F*	Sig.
Corrected model	312.39[a]	2	156.19	76.86	.00
Intercept	211.70	1	211.70	104.17	.00
Pre	1.75	1	1.75	.86	.35
Group	307.27	1	307.27	151.20	.00
Error	123.96	61	2.03		
Total	12619.00	64			
Corrected total	436.35	63			

[a] *R* squared = .716 (adjusted *R* squared = .707)

group appeared to have performed better on the oral proficiency post-test. A one-way ANCOVA test was performed in the following table to see whether there were any statistically significant differences between the two groups' oral proficiency post-test results:

There were significant differences between the experimental and control groups' oral proficiency post-test scores, as shown in Table 10, where Sig is.00, less than 0.05. On the post-test of oral proficiency, the experimental group students actually performed better than the control group students. It is plausible to assume that the advantages of using C-DA have contributed to the experimental group's improved performance on the oral proficiency post-test.

Discussion

The goal of the current study was to determine how two assessment methods—computerized dynamic assessment (C-DA) and rater-mediated assessment—affected the oral proficiency, writing performance, and test anxiety of EFL learners. The effectiveness of C-DA in reducing test anxiety and improving students' oral and written skills were both highly significant. In fact, the C-DA group fared better than the rater-mediated group. The study findings demonstrated how employing C-DA processes might significantly promote learners' performance. The development of students' writing and speaking skills and ability to lower their test anxiety through ongoing evaluation and self-modification via CDTW, which offers test-takers preplanned hints (mediation) incorporated in three steps of pre-writing, writing, and drafting and reformulation, can be used to describe the effectiveness of such an attitude.

Several significant conclusions were drawn after data analysis. In terms of speech proficiency and writing ability, C-DA groups considerably outperformed the rater-mediated group. This indicates that DA led to more complex, fluent, and accurate

spoken and written production. A plausible explanation is that students in DA groups were required to concentrate more on their oral and written productions than those in the rater-mediated group, who were required to focus on the grammar, length, and proper format of their written and oral productions as they were instructed. The velocity of their production and how to develop it along with the accuracy and complexity of their output were other factors that students considered because they were expected to deliver the most accurate, fluent, and complex responses (Ghahderijani et al., 2021).

The results of this study showed how effective C-DA was in improving students' oral communication skills and writing production. It revealed the role of C-DA in lowering pupils' test anxiety. When the results of this study were compared to those of other comparable studies, it became clear that this study's conclusions supported those of other studies, proving that DA is a generally effective method of language learning. These findings are consistent with other studies on DA that follows.

The outcomes are in line with study findings by Ghahderijani et al. (2021), which looked at the effects of two dynamic assessment (DA) models on speaking CAF. DA was seen as an interactive technique to assessment that combined teaching and assessment into a single instructional engagement as opposed to static assessment. Using ANOVA to analyze the data, it was discovered that C-DA and G-DA could both considerably boost speaking CAF compared to traditional non-DA training, with C-DA being significantly superior to G-DA.

These findings are also in agreement with study by Ebadi and Asakereh (2017), which looked at the development of EFL learners' speaking abilities using dynamic assessment. The participants narrated a series of visual stories for the data collection, receiving mediation based on their zone of proximal development (ZPD). To find any potential shifts in the participants' cognitive development, a paradigm for data analysis using microgenetic and theme analysis was used. The results showed that the individuals' cognition had significantly improved and that they were moving closer to full self-regulation.

Likewise, the research backs up the findings of the study by Moradian et al. (2019). He found that while the concurrent G-DA group in his study received calibrated feedback, the non-dynamic assessment (N-DA) group was explicitly given helpful assistance without taking into account their zone of proximal development (ZPD). The analysis of the data revealed that the G-DA group outperformed the N-DA group by a large margin. The efficiency of concurrent G-DA in learning requests and refusals was also revealed by the qualitative microgenetic analysis of the conversations between the students and their teachers, corroborating the effectiveness of dynamic assessment (DA) in pragmatic instruction.

Furthermore, these findings support Malmir's (2020) study, in which he investigated the effects of two models of dynamic assessment on the accuracy and quickness of perception of speech acts and implications. The results showed that, when compared to non-dynamic assessment teaching, both interactionist and interventionist models of dynamic assessment may considerably improve the Iranian EFL learners' pragmatic comprehension accuracy. This study also shown that interventionist dynamic assessment greatly increased pragmatic comprehension of requests, offers, suggestions, and speech acts as well as conventional and conversational implicatures.

In the same vein, the results of Ebrahimi's (2015) study, which found that using DA significantly enhanced speaking abilities, complexity, and accuracy, are also validated by this study. In her study, the control group got regular teaching according to the institute's standard procedure, whereas the experimental group received the intervention (DA). The results showed that DA implementation fostered more accurate and complex oral production while brought no impact on learners' fluency. Moreover, findings showed a considerable positive association between CAF measures and pupils' oral proficiency achievement.

The findings also support Ahmadi Safa et al. (2015) study results, which found that an interactionist model of DA had a statistically noteworthy positive influence on the speaking abilities of Iranian EFL pupils. The study's findings also confirm those of Talati-Baghsiahi and Khoshsima (2016), who examined the impact of a DA strategy on the linguistic and pragmatic knowledge of modal auxiliaries as hedging strategies among Iranian EFL students. They reasoned that the use of DA in EFL lessons enhanced pragmatic L2 language features like the provided hedges in writing activities.

These results can be attributed to specific characteristics of C-DA. The special characteristics that are present in DA models by definition account for the reliability of the C-DA model. The utilization of intense interaction between the intervener and the learner, which places the student at the center of all educational experiences, is the most crucial component of all dynamic models. Students are able to activate their existing knowledge and try to reach higher levels by obtaining scaffolding and aid from the instructor or other competent ones thanks to the extensive usage of interaction in DA lessons with the focus on the learning capacity of the learners (Lantolf & Poehner, 2011).

More significantly, ZPD, which involves an integrated teaching and evaluating procedure, is the main cause of DA general effectiveness. ZPD plays a crucial function and serves as the foundation of DA. The interaction between students and their teachers/assessors happens through DA and with ZPD in mind, and so activates the learners' potential to learn. To put it another way, assessing learning potential requires first identifying the ZPD before aiding the learner in recognizing and taking ownership of his own learning through interaction (Bekka, 2010).

Simply put, the goal of dynamic assessment use in the study was to offer learning-focused and growth-promoting challenges for an L2 learner in addition to evaluating and identifying what he currently knows. L2 learners can learn the target foreign language more successfully with the use of scaffolding within the zone of proximal development (ZPD), instructors' mediation, intensive engagement, and the conscientious participation (Kozulin & Garb, 2002).

Furthermore, as noted by Lantolf and Poehner (2011), the effectiveness of DA can be ascribed to students' increased exposure to and use of the target language in DA-focused courses. According to Kasper and Rose (2002), there is a connection between general learning and the volume of language exposure. Moreover, any DA-based interaction (whether with the first or second interactant) helped the learners to improve their cognitive performance and social involvement, as Poehner (2008) sharply said.

In addition to the specific characteristics of DA outlined above, the interest-provoking (González-Lloret, 2018) and encouraging qualities of C-DA may be used to explain how learners' progress in oral and written productions (Taguchi, 2019). González-Lloret

(2018) claims that doing so can improve learning by inspiring students, igniting their curiosity and imagination, removing the burden of traditional classrooms from their interactions, mixing inside- and outside-of-classroom learning, and eventually reducing their text anxiety.

Also, C-DA progresses gradually and offers the pupils opportunity to improve their competence. The complexity of cognitive processes that underlie neurological reactions in the brain can also be linked to the positive effects of the CDA model on speech proficiency and writing ability. Oral and written skills are the end product of these neurological mental processes, and three components of accuracy, complexity, and fluency may be easily identified in learners' performances (Taguchi, 2019).

Furthermore, it appears that the C-DA framework gives instructors some control over how well their students write or speak in a positive space by giving them learning opportunities that can help them become better communicators. This study also emphasizes the value of innovative learning environments that can help students regulate their learning strategies through C-DA and self-modification during the intervention phase, which enables them to learn more efficiently in a stimulating and challenging environment (Lee, 2010).

Besides, the fact that DA established a supportive environment to emphasize applicants' ongoing training and development by taking into account their ZPD may provide credence to the current findings. C-DA also offers a thorough diagnostic of the skills needed for mediators and learners in ZPD to actively intervene. This strategy switched the emphasis from the end result of prior learning to the processes through which abilities may be produced. With the provision of specific pre-determined cues and prompts as well as by assisting language instructors in projecting future performances, C-DA aids in the diagnosis of students' hidden learning issues. As a result, DA may create a better model of actual abilities and their development (Hidri & Roud, 2020).

Moreover, the learning environment is made more learner-friendly by including C-DA. Accordingly, C-DA lessens students' fears of failure, increases their enthusiasm for more study, and provides them the self-assurance they need to reach greater levels of functioning by demonstrating mastery of intervention supports (Ebadi & Saeedian, 2015). Most importantly, C-DA procedures, particularly, place a strong emphasis on both assessment and development of learners' skills, in contrast to typical psychometric assessment that only highlight the evaluation of learners' abilities. In this way, C-DA simply integrates instruction and evaluation, which necessitates awareness of learners' zones of proximal development (Zangoei et al., 2019).

Along with the rest, the ability to analyze a large number of students at once for how well they are achieving their potential is a clear benefit of adopting such tools. To disclose various patterns of learning abilities in challenging speaking and writing regions, DA is heavily dependent on meditational capabilities. The consequence is that learners can gain a lot from a DA-based mediation and that a C-DA-based intervention can have a significant part in the process of teaching L2 writing and speaking. DA describes a novel approach to teaching that is intended to produce notable educational results (Davoudi & Ataie-Tabar, 2015).

It can also be inferred that DA in general and C-DA in particular, aim to identify when pupils are having difficulty and making efforts to learn. According to Ajideh et al. (2012),

growing and developing student capacities is the ultimate goal of education, and this information aids educators in creating more effective courses by giving pertinent data regarding the origins of learner difficulties (Rashidi & Bahadori Nejad, 2018).

To sum up, C-DA, like the other DA methods, helps to attain the instruction-assessment integration aims and, in line with Vygotsky's theory, it revealed the learners' potential ability. On their ZPD, this was accomplished. The results of C-DA assessments provide access to go much beyond those of non-dynamic assessments, which ignore the individual learning styles of students. To put it another way, it is not always accurate to assume that two learners who perform equally on a pretest also do equally well. If this is the case, it would be impossible to argue that a participant would outperform another in terms of having a greater variety of potential if they did not participate in the C-DA. To put it another way, their learning potentials didn't exhibit minor differences until the C-DA was used to determine how much mediation was required for each ability (Ebadi & Saeedian, 2015).

Conclusion and implications of the research

This investigation attempted to comparatively inspect the impacts of rater-mediated and C-DA on EFL learners' writing performance, test anxiety, and oral proficiency. The results illustrated that using these two kinds of assessments enhanced EFL learners' writing performance, oral proficiency, and lowered their test anxiety. C-DA group showed superior improvement than the rater mediated assessment group on post-tests.

In other words, pupils were able to generate more fluent, complex, and accurate answers, both in writing and oral skills outputs. In summary, it can be said that the distinctive characteristics built into DA models may be credited for the development in writing performance and speaking competence abilities. According to Lantolf and Poehner (2011), the usage of extensive interactions between both the learners and interveners, which places the learner at the focus of whole educational practices, is the most crucial component of all dynamic models.

The extensive application of interactions in DA courses, with an emphasis on the learning potential of the students, permits the students to make use of their existing knowledge and work toward higher levels by getting scaffolding and support from the teacher or other competent individuals (Lantolf & Poehner, 2011). Instruction and assessment are blended dialectically in DA, as recommended by Vygotsky's ZPD, to guide learners toward a future that is always evolving and dynamic.

Additionally, the effectiveness of DA can be linked to students' increased use and exposure to the target language in DA-focused courses (Lantolf & Poehner, 2011). Besides, the quantity of language contacts and general learning are directly related (Kasper & Rose, 2002). Also, any DA-based link aids students in enhancing their social engagement and cognitive performance (Poehner, 2008). According to Tajeddin and Tayebipour (2012), ZPD-based interactions support the supremacy of DA over N-DA models by claiming that interactions in student's ZPD offer a chance for having mastery in a variety of skills.

The specific characteristics of DA outlined above as well as the interest-rising and motivational qualities of computer-based instruction can also be used to explain the substantial influence of C-DA in oral competence compared to the traditional N-DA

(Taguchi, 2019). As a result, C-DA can improve learning by inspiring L2 students, igniting their curiosity and creativity, creating a less intimidating setting for exchanges free from the pressure of traditional classrooms (González-Lloret, 2018).

This study can have several implications for curriculum developers, educators, and students. with the help of EFL teachers, language learners become exposed to a range of assessments, including the topic of the present study. Through the use of assessment-based mediation, teachers may better support pupils. Educators may find it helpful to utilize these assessments as they are useful instruments to develop students' speaking and writing abilities. Students' test anxiety can be decreased by engaging them in activities in the classroom as well as employing a range of assessments. EFL learners can gradually realize how to be self-directed learners as a consequence. The current research offers a strong basis for incorporating information about basic features of oral and written productions into teaching and assessment procedures and using it to advance learners' performance.

For EFL students, the results may suggest that familiarity with various assessment types, especially C-DA and rater-mediated assessment, can help them improve their spoken and written skills. Additionally, learners may identify specifically where they want support and aid so that they can approach those who are qualified for assistance. Additionally, since this study found that C-DA had positive impacts on EFL students' speaking and writing abilities, it is possible to develop teaching strategies and course materials for language classes that promote computerized techniques and subsequently affect the language abilities of foreign language learners. Syllabi creators are urged to include various assessment modalities in their instructions. The study's conclusions may also be used by content creators to create a variety of activities and tasks that are appropriate for L2 learners of varying skill levels.

This study, like other studies, had limitations and couldn't address and cover all pertinent issues. This study's flaws were its inability to accurately characterize students' achievement using a computer scoring method. The current study's intervention period was also quite brief. Additionally, the pupils' long-term learning was not assessed. When the students started to demonstrate some improvement on the activities, the data gathering process came to an end. Last but not least, the study's design was to evaluate a significant number of students' written and spoken productions using just an interventionist technique. The questions in the current study were in limited format.

The research also only included those who fell inside a certain age range. The results cannot thus be applied to other age categories. There was a maximum of 64 participants in this research. Therefore, this cannot be used broadly. The study only included male students, so it is plausible that the results do not generalize to female students.

A few recommendations are provided for further studies. Future research might broaden the focus to an extended time period in numerous sessions and assess the impacts of C-DA on boosting learner improvement, that can be extremely important to find the proof of transfer as the students require time to internalize what they have learnt in the process. Therefore, further research to monitor writing and speaking development over longer period of teaching would offer more insights.

It is advised that further researchers create and deploy C-DA programs of a similar nature. In this regard, a variety of empirical investigations using various meditational

approaches would be required to uncover various learning patterns in the challenging domains of speaking and writing. Large-scale assessment will need to incorporate more techniques in addition to interactionist dynamic process, and the C-DA technique will need to be applied to different formats.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

Ahmadi Safa, M., Donyaie, S., & Malek Mohammadi, R. (2015). An investigation into the effect of interactionist versus interventionist models of dynamic assessment on Iranian EFL learners' speaking skill proficiency. *Teaching English Language*, *9*(2), 146–166.

Ajideh, P., Farrokhi, F., & Nourdad, N. (2012). Dynamic assessment of EFL reading: Revealing hidden aspects at different proficiency levels. *World Journal of Education*, *2*, 102–111.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*, (vol. 1). Oxford University Press.

Bailey, K. M., & Brown, J. D. (1984). A categorical instrument for scoring second language writing skills. *Language Learning*, *34*(4), 21–38.

Bakhoda, I., & Shabani, K. (2016). 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*(1), 1–16.

Bekka, K. G. (2010). Dynamic assessment for learning potential: A shift in the focus and practice of evaluating Japanese oral proficiency. *Japanese Journal of Education*, *10*(1), 53–66.

Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers: A research based guide to instructional programs and practices*. International Reading Association.

Breland, H. M., Bonner, M. W., & Kubota, M. Y. (1995). Factors in performance on brief impromptu essay examinations. *ETS Research Report Series*, *1995*(2), i–35.

Carney, J. J., & Cioffi, G. (1990). Extending traditional diagnosis: The dynamic assessment of reading abilities. *Reading Psychology: An International Quarterly*, *11*(3), 177–192.

Chodorow, M., & Burstein, J. (2004). Beyond essay length: Evaluating e-rater®'s performance on toefl® essays. *ETS Research Report Series*, *2004*(1), i–38.

Craig, K. J. (1995). Environmental factors in the etiology of anxiety. In *Psychopharmacology: The fourth generation of progress*, (pp. 1325–1339).

Cronbach, L. I. (1990). *Essentials of psychological testing*, (5th ed.,). Harper and Row.

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.

Ebadi, S., & Asakereh, A. (2017). Developing EFL learners' speaking skills through dynamic assessment: A case of a beginner and an advanced learner. *Cogent Education*, *4*(1), 1419796. https://doi.org/10.1080/2331186X.2017.1419796.

Ebadi, S., & Saeedian, A. (2015). The effects of computerized dynamic assessment on promoting at-risk advanced Iranian EFL students' reading skills. *Issues in Language Teaching*, *4*(2), 26–21.

Ebrahimi, E. (2015). The effect of dynamic assessment on complexity, accuracy, and fluency in EFL learners' oral production. *International Journal of Research Studies in Language Learning*, *4*(3), 107–123.

Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*, (2nd ed.,). Peter Lang.

Engelhard, G., & Wind, S. A. (2017). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge.

Engelhard Jr., G., & Wind, S. A. (2019). Introduction to the special issue on rater-mediated assessments. *Journal of Educational Measurement*, *56 3)*, 475–477.

Esfandiari, R. (2021). Rater-mediated assessment of iranian undergraduate students' college essays: Many-facet rasch modelling. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, *9*(1), 93–119.

Genesee, F., Lindholm-Leary, K., Saunders, W. M., & Christian, D. (2006). *Educating English language learners*. Cambridge University Press.

Ghahderijani, B. H., Namaziandost, E., Tavakoli, M., Kumar, T., & Magizov, R. (2021). The comparative effect of group dynamic assessment (GDA) and computerized dynamic assessment (C-DA) on Iranian upper-intermediate EFL learners' speaking complexity, accuracy, and fluency (CAF). *Language Testing in Asia*, *11*(1), 1–20.

González-Lloret, M. (2018). Pragmatics in technology-mediated contexts. In A. Herraiz-Martínez, & A. Sánchez-Hernández (Eds.), *Learning second language pragmatics beyond traditional contexts*, (pp. 15–46). Peter Lang.

Harmer, J. (2006). *The practice of English language teaching*, (8th ed.,). Longman.

Hasson, N., & Botting, N. (2010). Dynamic assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy*, *26*(3), 249–272.

Heritage, M. (2012). From formative assessment: Improving teaching and learning. In *Paper presented at the CRESST 2007 Assessment Conference, Los Angeles, CA*.

Hidri, S., & Roud, L. F. P. (2020). Developing and using hints in computerized dynamic assessment of a TOEFL iBT reading exam. *Heliyon*, *6*(9), e04985.

Iwashita, N. (2010). Features of oral proficiency in task performance by EFL and JFL learners. In *Selected proceedings of the 2008 second language research forum*, (pp. 32–47). Cascadilla Proceedings Project.

Kasper, G., & Rose, K. R. (2002). Pragmatic development in a second language. In *Language learning*.

Khodashenas, M. R., & Rakhshi, F. (2017). The effect of electronic portfolio assessment on the writing performance of Iranian EFL learners. *International Journal of Research in English Education*, *2*(3), 67–77.

Knoch, U., Zhang, B. Y., Elder, C., Flynn, F., Huisman, A., Woodward-Kron, R., … McNamara, T. (2020). I will go to my grave fighting for grammar: Exploring the ability of language-trained raters to implement a professionally-relevant rating scale for writing. *Assessing Writing*, *46*, 1–14. https://doi.org/10.1016/j.asw.2020.100488.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, *16*(3), 154–169.

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, *23*(1), 112–127.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, *1*(2), 49–72.

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, *15*(1), 11–33. https://doi.org/10.1177/1362168810383328.

Laufer, B., & Hill, M. (2000). *What lexical information do L2 learners select in a CALL Dictionary and how does it affect word retention?*

Lee, Z. H. (2010). *An experimental study on situated and dynamic learning assessment (SDLA) environment*. The University of North Texas.

Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In *Vygotsky's educational theory in cultural context*, (pp. 99–116).

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

Madsen, H. S. (1983). *Techniques in testing*. Oxford University Press.

Mallahi, O., & Saadat, M. (2020). Effects of feedback on Iranian EFL learners' writing development: Group dynamic assessment vs. formative assessment. *Iranian Evolutionary and Educational Psychology Journal*, *2*(4), 258–277.

Malmir, A. (2020). The effect of interactionist vs. interventionist models of dynamic assessment on L2 learners' pragmatic comprehension accuracy and speed. *Issues in Language Teaching*, *9*(1), 279–320.

McNamara, T. F. (1996). *Measuring second language performance*. Addison Wesley Longman.

Moradian, M., Asadi, M., & Azadbakht, Z. (2019). Effects of concurrent group dynamic assessment on Iranian EFL learners' pragmatic competence: A case of requests and refusals. *Journal of Research in Applied Linguistics*, *10*(2), 106–135.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, *4*(4), 386–422.

Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, *5*(2), 189–227.

Ngoc Anh, D. T. (2019). EFL student's writing skills: Challenges and remedies. *Journal of Research and Method in Education (IOSR-JRME)*, *9*(6), 74–84.

Nunan, D. (2003). *Practical english language teaching (international edition)*. McGraw-Hill.

Pappamihiel, N. E. (2002). English as a second language students and English language anxiety: Issues in the mainstream classroom. *ProQuest Education Journal*, *36*(3), 327–355.

Pishghadam, R., & Barabadi, E. (2012). Constructing and validating computerized dynamic assessment of l2 reading comprehension. *Iranian Journal of Applied Linguistics (IJAL)*, *15*(1), 73–95.

Pishghadam, R., Barabadi, E., & Kamrood, A. M. (2011). The differing effect of computerized dynamic assessment of L2 reading comprehension on high and low achievers. *Journal of Language Teaching and Research*, *2*(6), 1353–1358.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Springer. https://doi.org/10.1007/978-0-387-75775-9.

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, *9*(3), 233–265.

Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment. *Language Teaching Research*, *17*(3), 323–342.

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.

Popham, W. J. (2008). *Classroom assessment: What teachers need to know*, (5th ed.,). Prentice Hall.

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), 2158244018784643.

Richard, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics*, (3rd ed.,). Longman.

Robinowitz, A. (2010). From principles to practice: an embedded assessment system. *Applied Measurement in Education*, *13*(2), 181–208.

Sarason, I. G. (1988). Anxiety, self-preoccupation and attention. *Anxiety Research*, *1*(1), 3–7.

Sepehrian, A. (2013). Self-Efficacy, achievement motivation and academic procrastination as predictors of academic achievement in pre-college students. *Proceeding of the Global Summit on Education*, *6*, 173–178.

Shabani, K. (2021). Diagnostic and developmental potentials of computerized dynamic assessment (C-DA) for L2 vocabulary. *Interdisciplinary Studies in English Language Teaching*, *1*(2), 165–187.

Shokrpour, N., & Fallazadeh, M. (2007). A survey of the students and interns EFL writing problems in Shiraz university of medical sciences. *Asian EFL Journal*, *9*(1), 77–89.

Sieber, J. E. (1980). Defining test anxiety: Problems and approaches. In *Test anxiety: Theory, research, and applications*, (pp. 15–40).

Spielberger, C. D. (1972). Anxiety as an emotional state. In *Anxiety-Current trends and theory*, (pp. 3–20).

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing*. Cambridge University Press.

Taguchi, N. (2019). Comprehension of conversational implicature in L2 Chinese. *Pragmatics & Cognition*, *21*(1), 139–157. https://doi.org/10.1075/pc.21.1.06tag.

Tajeddin, Z., & Tayebipour, F. (2012). The effect of dynamic assessment on EFL learners' acquisition of request and apology. *The Journal of Teaching Language Skills (JTLS)*, *4*(2), 88–118.

Talati-Baghsiahi, A., & Khoshsima, H. (2016). Improving linguistic and pragmatic knowledge of hedging strategies in EFL undergraduate students: a dynamic assessment approach. *International, Journal of English Language & Translation Studies*, *4*(2), 13–28.

Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning*, *18*(1), 21–32.

Valasa L, Mason LH, Benedek-Wood E. (2009). Teaching low-achieving students to self-regulate persuasive quick write responses. *Journal of Adolescent and Adult Literacy*, *53*(4), 302-12.

Van der Veen, C., Dobber, M., & van Oers, B. (2016). Implementing dynamic assessment of vocabulary development as a trialogical learning process: A practice of teacher support in primary education schools. *Language Assessment Quarterly*, *13*(4), 329–340.

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

Vygotsky, L. S. (1986). *Thought and language-revised edition*. Massachusetts Institute of Technology.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Wind, S. A. (2020). Do raters use rating scale categories consistently across analytic rubric domains in writing assessment? *Assessing Writing*, *43*, 100416.

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress & Coping*, *17*(3), 227–240.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53–67.

Yang Y, Qian DD. (2019). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, 1–25. https://doi.org/10.1080/09588221.2019.1585882

Zangoei, A., Zareian, G., Adel, S. M. R., & Amirian, S. M. R. (2019). The impact of computerized dynamic assessment on Iranian EFL learners' interlanguage pragmatic development. *Journal of Modern Research in English Language Studies*, *6*(4), 165–139.

Zhang, J., & Lu, X. (2019). Measuring and supporting second language development using computerized dynamic assessment. *Language and Sociocultural Theory*, *6*(1), 92–115.

Publisher's Note