An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment

Saman Ebadi*, Elham Karimi and Shokoufeh Vakili

*Correspondence:
samanebadi@gmail.com

Applied Linguistics, Razi
University, Kermanshah, Iran

Abstract

As a part of a more extensive computerized dynamic assessment (CDA) project delivered through a dedicated website, www.lingeli.com, this study explores EFL learners' perspectives on an online listening comprehension dynamic assessment software focusing on inferential listening skills while trying to minimize the guessing effect. The study participants were 94 Iranian EFL learners selected through convenience sampling to participate in a 1-month dynamic assessment targeting EFL learners listening comprehension using the developed software. The researchers involved half of the study participants in semi-structured interviews held in their native language (Persian) to shed light on the pros and cons of the software. Six major themes emerged from the interviews' content analysis, encapsulating participant" generally positive perceptions and some negative perspectives about the software. They perceived the CDA software as novel and interesting, stress-relieving, supportive, convenient, and cost-effective in terms of time and money, and removing the time and location limitations. In contrast, a few others assumed it was time-consuming and stressful, with limited learning and ignoring the benefits of social aspects of teaching. The study's findings recommend that learners' perspectives be considered when developing future CDA listening comprehension software to increase the validity of CDA projects by addressing their drawbacks and fully implementing the CDA test's potential to enhance EFL learners' listening comprehension abilities.

Keywords: Computerized dynamic assessment, Inferential listening comprehension skills, Listening comprehension skills, Learners' perspectives, Qualitative research

Introduction

Computerized dynamic assessment (CDA), as a new strand of dynamic assessment (DA), originated from the Russian psychologist Lev Vygotsky's Sociocultural Theory (SCT), relying on the reintegration of assessment and instruction to simultaneously account for the developed and the developing capabilities of individuals (Sternberg & Grigorenko, 2001). CDA enjoys an interventionist approach to providing automatic pre-scripted mediation to learners through computers. In this way, CDA has enabled researchers to target not only second/foreign language (L2/FL) skills in general (e.g., Ebadi et al., 2018; Ebadi & Saeedian, 2015; Estaji & Saeedian, 2020; Li & Zhu, 2013, 2017; Ma, 2017; Mehri

Kamrood et al., 2019; Teo, 2012; Yang & Qian, 2017, 2019), but also listening comprehension skill which has been regarded as the most challenging language learning skill for EFL learners to overcome (Kern, 2000).

In our comprehensive project, CDA was applied to targeting inferential listening comprehension skills as one of the main mechanisms involved in producing or understanding spoken or written discourse (Kern, 2000). Ableeva (2010) has argued that listening comprehension is an inferential mechanism that requires working on the details that are not always obviously provided in the text. Inferential comprehension requires listeners to integrate concepts, make inferences, analyze and assess information, and recognize tone and voice. On the other hand, the software aimed at minimizing the guessing effect. Alnabhan (2002) regards guessing as one of the primary mechanisms that negatively affect listening comprehension skills. Thus, negative marking could effectively minimize pure guesswork and contribute to more accurate test scores (Jaradat & Sawaged, 1986). Increasing the number of options in multiple-choice queries can also be used as the main strategy of test developers to minimize the impact of guessing (Burton, 2001, 2004; Karandikar, 2010; Mameren & Vleuten, 1999; Zimmerman & Williams, 2003). However, Lesage et al. (2013) state that it is quite probable that additional options do not serve as productive distractors and miss their function. Also, there is no guarantee that the test taker has not guessed the correct response, resulting in a false reporting of actual and mediated scores. Therefore, our CDA software introduced guess-checking questions as a new strategy for controlling the guessing effect in assessment.

The CDA test software, delivered through the dedicated website, www.lingeli.com, acted like an exam and provided users with gradually delivered prompts after each failure in choosing the right answers. More precisely, the interventionist mediation was applied in a cake format (Sternberg & Grigorenko, 2002) to assess and trigger learners' development using the web-based multiple-choice CDA software developed by the researchers. It is noteworthy that, in developing the CDA software, the researchers followed the steps in the listening section of Poehner and Lantolf's (Poehner & Lantolf, 2013) CDA project designed for languages other than English. Thus, our CDA software provided us with quantitative data on learners' performance concerning the targeted features (inferential listening comprehension and guess-checking).

Vandergrift (2007) notes that in order to obtain insights into the mechanisms that cannot be touched upon through quantitative methods, L2 listening studies can primarily implement qualitative methods such as interviews to determine what students are attending and why they are paying selective attention to it. For this reason, in the present research, a semi-structured interview was used to conduct diagnostic and process-based research in L2 listening comprehension by exploring participants' attitudes toward the newly developed CDA software. In sum, this study was guided by the following question;

RQ: What are Iranian EFL learners' perspectives toward using the inferential listening comprehension CDA software?

Literature review

The influence of DA on listening skills has been highlighted over the last few decades. Ableeva (2008) applied DA to French listening comprehension and concluded that, unlike the case with the non-dynamic pre-test, the DA approach uncovered the unique

development capacities of the learners. Similarly, Shabani (2014) investigated the effects of DA on L2 learners' listening comprehension and found that DA increased and eased college L2 learners' listening comprehension. The results also proposed that applying the DA approach to listening tasks may help educators in language classrooms recognize comprehension issues. The influence of non-DA and DA treatment on listening comprehension of 60 Tunisian L2 college students was also compared by Hidri (2014), who confirmed that the DA approach increased students' cognitive development and resulted in substantial gains in listening comprehension compared to the traditional assessment techniques.

The DA contributions to Chinese EFL learners' listening comprehension were examined by Wang (2015). The researcher selected five subjects and played audiotapes for them during the pre-test sessions. After recognizing their difficulties, the researcher mediated the students and replicated their learning sessions to assess their degree of assimilation. The study found that DA could provide a more in-depth analysis of the students' listening comprehension problems. In another research, the schema theory's effectiveness on students' listening skills was examined by Farangi and Saadi (2017). The researchers chose two groups of learners, each of them including 42 members. The DA group participated in a pretest-enrichment-posttest procedure, whereas the schema theory group went through a pre-listening, listening, and post-listening procedure, including shadowing and semantic mapping. The results showed that although both groups established and strengthened their listening skills, there was no significant difference between their development.

Using Vygotsky's (1978) socio-cultural theory as a conceptual foundation, Al-Dawoody Abdulaal et al. (2022) aimed to evaluate the effect of dynamic assessment (DA) vs. non-dynamic assessment (NDA) on Ethiopian intermediate EFL learners' receptive skills. This study involved 96 intermediate learners from a high school. They were then split into three equally sized groups: one control group (CG) and two experimental groups (EG1 and EG2). After giving a pretest, the EG1 and EG2 received group dynamic assessment-based training, whereas the control group received conventional instruction. A post-test was conducted after the treatment. One-way ANCOVA analysis showed that dynamic assessment significantly affected receptive skills. This study had implications for educators, learners, and material designers. Teachers are recommended to incorporate DA into their language lessons to assist pupils with strengthening their English language skills.

In spite of the increasing interest in adapting the DA approach to computer-based EFL/L2 studies in the past few decades (e.g., Darhower, 2014; Ebadi et al., 2018; Ebadi & Bashir, 2020; Ebadi & Rahimi, 2019; Ebadi & Saeedian, 2019; Estaji & Saeedian, 2020; Karlström & Lundin, 2013; Li & Zhu, 2017; Ma, 2017; Mehri Kamrood et al., 2019; Oskoz, 2005; Vakili & Ebadi, 2019; Yang & Qian, 2019) only a few studies have targeted listening comprehension skills. Poehner et al. (2014) investigated the improvement of reading and listening comprehension skills through an online electronic mediation program. Although the online mediation in their study was an innovation, they also used a fixed and standardized mediation design identical to the previous studies (e.g., Ableeva, 2008). They utilized multiple-choice online assessment, gradually offering mediation from the most implicit to the most explicit prompts. They made the diagnosis based

on the test-takers' actual and mediated scores. According to the study's findings, these two scores, together with an analysis of the test-takers' performance on particular items, provided a precise diagnostic of their L2 progress and useful data for further instruction and learning.

Ebadi and Vakili Latif (2015) investigated the efficiency of electronic concept mapping in enhancing EFL learners' listening skills by examining relevant changes in their ZPDs. After each listening track in ten DA sessions, the participants were instructed to make concept maps. According to the study's findings, the participants' ability to go from the feedback intervention stage to self-regulated learning demonstrated that their listening comprehension ZPDs had developed. In another attempt, Mashhadi Heidar and Afghari (2015) conducted web-based research to examine how dynamic listening comprehension assessment could help EFL learners develop their socio-cognitive abilities. They emphasized the actual potential of individual students and the evaluative potential of the developmental phase of listening.

Drawing on previously created software by Poehner and Lantolf (2013), Mehri Kamrood et al. (2019) constructed and executed an interventionist online computerized program for the dynamic assessment of the listening abilities of Iranian EFL students. The program offered students ZPD-based mediation through hints and prompts. It generated three scores: an actual score, a mediated score, and a Learning Potential Score (LPS), which reflected the difference between the actual and mediated scores. The results showed a significant difference between learners' actual and mediated scores, illustrating the insufficiency of non-dynamic evaluation in taking learners' responsiveness to mediation into account. Additionally, LPS could distinguish between students placed at the same level by non-dynamic assessment. It was also helpful to look at each learner's scoring profile and LPS to diagnose their strengths and weaknesses in key language constructs and develop carefully catered, tailored lesson plans and resources for their future learning.

An online computerized test of implicature comprehension (i.e., indirect meaning) was designed and administered by Qin & van Compernolle, 2021. The assessment's primary focus was on indirect acceptances, refusals, and opinions. Its dual objectives were to examine test-takers' (a) existing abilities as shown by their independent performance and (b) developing abilities as shown by their responsiveness to mediation. Sixty-seven individuals studying Chinese as a second language at US universities, evenly divided into beginner, intermediate, and advanced levels, took the test. Despite the fact that all test takers improved after mediation, the results showed that there was a large amount of individual variation in terms of response to mediation. The findings demonstrated that performance that underwent mediation was more instructive as a diagnosis of the learner's capacities compared to autonomous performance alone.

Kao and Kuo (2021) applied CDA principles using iSpring Quiz Maker to evaluate the unique learning requirements of 172 L2 English learners, diagnose their listening difficulties, and optimize their potential for future learning. The junior high school students who took part in this CDA program had their performance evaluated on three different types of English listening questions—overview, detail, and inference questions. The test results showed each participant's present performance, mediated performance, learning needs, and learning potential score. The study findings revealed that (1) participants

encountered challenges while responding to all three types of questions, (2) the CDA program's mediation was most effective in assisting participants in responding to inference questions, and (3) participants with low LPSs had more difficulty answering inference questions while participants with mid- or high-level LPSs needed more assistance when answering detail questions. Even when they earned the same scores, the participants' listening difficulties for the three different question types varied considerably. The study illustrated how CDA might give a comprehensive picture of EFL students' English listening skills, which could therefore have an impact on English pedagogical strategies and instructional design.

The effects of two dynamic assessment models on the speed and accuracy of L2 pragmatic comprehension were examined by Malmir and Mazloom (2021). The study included a convenience sample of 52 upper-intermediate female EFL learners who were randomly assigned to three groups: two experimental groups of dynamic assessment (GDA), a computerized dynamic assessment (CDA), and a non-DA control group. The pre- and post-test was a 26-item pragmatic listening comprehension test created by the researchers, and 14 sessions of treatment were conducted. The findings showed that CDA and GDA may dramatically improve pragmatic comprehension accuracy when compared to conventional non-DA training, with CDA being noticeably superior to GDA. According to the study's findings, using CDA in the classroom can improve pragmatic learners' comprehension speed and accuracy.

Pileh Roud and Hidri investigated the effects of computerized dynamic assessment (CDA) on the TOEFL iBT listening comprehension test administered to 185 Iranian EFL students who enrolled in TOEFL exam preparation classes at various language centres in Iran (Pileh Roud & Hidri, 2021). In order to generate actual, mediated, and learning potential scores, test-takers were given hints to help them comprehend the listening questions. The results of the study demonstrated that, for almost all question types, the mean scores created by the actual and mediated scores varied significantly across different levels of listening comprehension ability. Furthermore, the findings showed that CDA considerably enhanced EFL test-takers' performances in conducting the monologue and dialogue tasks.

Moreover, some researchers focused on learners' perspectives on DA and CDA procedures. For instance, Modarresi and Alavi (2014) investigated how DA and non-DA learners perceived CDA and discovered the following common perceptions: more opportunities (19.1%), novelty (20.7%), reaction to scores (8%), feeling at ease (15.4%), motivation (6.8%), and test timing (5.9%). The study's findings indicate that when CDA is used, students experience an emotionally safe setting, less stress and anxiety, less worry about the consequences of their failure, and a boost in self-esteem and self-concept. In another study, Ebadi and Yari (2015), by conducting a post-study semi-structured interview and using thematic analysis on the interview transcripts, sought to learn about learners' perceptions of the efficacy of DA mediation on vocabulary development. The study's results indicated that the participants had positive viewpoints toward implementing DA in instructing vocabulary.

Adokh and Rafiee (2017) made another effort to explain the procedure and applicability of DA based on the viewpoints of EFL students and teachers in the Iranian educational system. After offering an interventionist DA model in which participants were

instructed to compose essays, 25 EFL teachers and 45 EFL students were required to complete two questionnaires. The findings showed that although their opinions were not entirely positive, student participants had positive perceptions of DA in language schools. At the end of another CDA study on reading comprehension conducted by Yang and Qian (2017), a questionnaire was used to gather information about learners' opinions on CDA. According to the survey's findings, CDA was warmly welcomed by the students and was thought to be especially beneficial in boosting the self-confidence of those who performed poorly on standard static assessments.

Babamoradi et al. (2018) also investigated the learners' attitudes regarding the instruction and assessment of writing through CDA. The researchers created software that included 11 tests with multiple-choice questions. Throughout each test, participants received feedback depending on their requirements and Zone of Proximal Development (ZPD). The students were expected to record their opinions regarding the efficacy of CDA in teaching and assessing writing after each lesson and test. After reviewing their diaries, it was discovered that every learner had a positive view of the use of CDA in writing instruction. As the final step of another study, Ebadi and Rahimi (2019) conducted a semi-structured interview to explore the learners' opinions on the effects of online DA on their academic writing skills. The learners' positive perceptions of the impacts of online DA on improving online DA's impact were underlined by thematic analysis, which was utilized to analyze the data gathered through interviews.

As the reviewed studies imply, numerous studies have demonstrated the effectiveness of DA and CDA procedures in the instruction of listening comprehension. They have noted positive attitudes regarding applying this assessment in teaching different components of English. Additionally, there is a gap in the literature about reflecting learners' perspectives on the ways and means to target listening comprehension skills. These encouraging findings prompted us to learn about learners' perspectives on utilizing CDA in teaching and assessing inferential listening comprehension skills.

Method

Study design

This study (part of a bigger project) adopted a multiple case study design and undertook semi-structured interviews to investigate learners' perspectives on an online computerized dynamic assessment of inferential listening comprehension skills. According to Lodico et al. (2010), a case study is a type of qualitative research that aims to gain an in-depth understanding of an individual, group, or circumstance. In a multiple case study, different cases are examined to identify their similarities and differences. According to Baxter and Jack (2015), this allows the researcher to analyze both within and between settings. When cases might not accurately represent the entire population they were chosen from, multiple cases are sometimes preferred to a single case. However, choosing several cases restricts how deeply each case may be examined. According to Yin (2003), multiple case study approaches are more reliable since they offer more evidence from various sources. Although Baxter and Jack (2015) note that data gathered from several case studies is strong and reliable, they also note that gathering such evidence can be very time-consuming and expensive.

In the project's first stage, the researchers developed a web-based multiple-choice CDA software to assess and promote Iranian EFL learners' inferential listening comprehension skills. This was done by applying the interventionist approach of DA in a cake format (Sternberg & Grigorenko, 2002) using the listening section of Poehner and Lantolf's (Poehner & Lantolf, 2013) online CDA project designed for languages other than English. After collecting the data, the researchers fed the quantitative data into SPSS to carry out the analytical process. Two paired-sample *t* tests were run to compare EFL learners' mediated and actual scores to explore the development of their inferential listening and compare the actual scores of learners with and without calculating guessing effect scores, respectively. The results showed that learners' responsiveness to mediation significantly changed their performance on the CDA of inferential listening skills. Additionally, adding guess-checking items significantly decreased the impact of guessing in the CDA multiple-choice tests, which provided the opportunity to obtain a more comprehensive picture of learners' abilities. The findings provided more insights into the effective role of CDA in improving EFL learners' inferential listening comprehension skills. They called for the implementation of CDA programs to diagnose learners' deficiencies and strengths and uncover their zones of proximal development.

Afterward, to investigate learners' perspectives, the study conducted volunteer semi-structured individual interviews (Drever, 1995) with half of the initial participants to identify the strong and weak points of the software. A review and analysis of the literature that came before the interviews served as a design manual for them.

Richards (2009) points out that the dialogic nature of the interview offers significant evidence for exploring individuals' attitudes toward the data analysis process and highlighting underpinning values, which necessitate more flexibility for in-depth investigation of its subtleties. Furthermore, the qualitative approach reveals truth through how people perceive their surroundings (Merriam & Tisdell, 2015). As previously indicated, this study explores how learners perceive using CDA in teaching and assessing inferential listening comprehension skills, and this renders personal interviews the most appropriate tool.

In this phase, open-ended questions were used in 47 semi-structured interviews to allow participants to discuss their experiences, perceptions, and concerns. According to Stuckey (2013), semi-structured interviews let the researcher create an outline of the themes that will be covered. However, because it is frequently preceded by observation or informal interviewing, the interview is guided by the interviewee's responses. This aids researchers in better comprehending the topic and generating appropriate interview questions (Stuckey, 2013). The interviews were then coded for transcription. Following that, content analysis was conducted to reveal the major themes in the data.

Participants and sampling

The selection and number of participants are determined by the research objectives and the characteristics of the study population (Mack & Woodsong, 2005). This study needed samples similar in terms of their initial inferential listening skill. Therefore, it used a convenience sampling method and worked with 94 (38 males and 56 females) Iranian EFL university students for the first stage, wherein the learners were engaged in using the software. DonYei (2007) argues that convenience sampling fulfills specific

practical requirements such as ease of access or similarity in proficiency level as required in any given study. For the second stage of the study, where learners' perspectives were targeted, only volunteer learners were involved in the interview process. Given that, 47 study participants were interviewed, and their ideas were reflected in this paper.

Ethical considerations in this study were observed by ensuring participants that their collaboration was voluntary and that the data from their performance would remain confidential and anonymous. Moreover, the participants were assigned numbers P1, P2, ..., and P94 to protect their identities throughout the research.

Data collection

A total of 108 Iranian EFL learners (47 men and 61 women) with a B2 proficiency level reported by DIALANG, took the CDA test at first. The participants in the study were briefed about the CDA project and were made aware of the nature of the research. Following that, students were given a web link to participate in the 80-min CDA test of inferential listening comprehension skills available at www.lingeli.com . Ninety-four original volunteers (38 men and 56 women) were chosen to participate in the study. The remaining students were omitted since the CDA test analysis showed that they did not devote enough time to it. After passing the test, individuals could generate their scoring profiles on the website.

Finally, semi-structured individual interviews were performed to explore participants' attitudes on the effects of the CDA software on their inferential listening comprehension skills. The researcher interviewed 47 students, which constituted half of the study participants. Each participant was interviewed for around 15 min in their mother tongue (Persian) to thoroughly minimize confusion and elicit learners' perceptions on the issue. The interviews were audiotaped, transcribed, and translated into English to conduct content analysis. To preserve learners' privacy, their consent was requested to record their voices and use the data for research purposes.

The semi-structured interviews included questions about learners' perceptions of the usefulness of the CDA software employed in this study. Participants' personal preferences and experiences throughout the test and suggestions for more effective CDA implementation were also investigated. Finally, the interviews ended by asking the participants about any concerns not addressed in the interview questions.

Instruments

The following instruments were used in different phases of our study:

1. A web-based CDA software

The current study's researchers designed a web-based program to assess and improve the inferential listening comprehension skills of Iranian EFL learners. The program replicates the listening part of Poehner and Lantolf's (Poehner & Lantolf, 2013) online CDA project that was created for languages other than English (with slight modifications). The ASP.NET programming language was used to create the software. There stages of test development, test piloting, and software development were followed in the CDA software development process.

Twenty multiple-choice items, followed by a short audio recording lasting about 30 s, made up our CDA exam. Lantolf and Poehner (2013) used a multiple-choice format for the test items in their CDA study. Each question had five options, so there would be enough room for each option to be attempted multiple times without immediately revealing the right choice. This study follows the same format: a multiple-choice test with five options for each question. In what follows, a test item is exemplified (Fig. 1).

Four prompts were set for each question following Aljaafreh and Lantolf's (Aljaafreh & Lantolf, 1994) regulatory scale for providing ZPD-based feedback. This study presented its ZPD-based prompts like Aljaafreh and Lantolf (1994); they were conditional (provided only when test-takers requested them) and graduated (organized from the most implicit to the most explicit level). In this approach, hints 1 and 4 were the same for all items since the first one, the most implicit hint, invited learners to try the item again after failing to select the correct answer on their first try.

The latter was the correct answer, followed by its explanation of the most explicit hint. However, depending on the nature of the item, hints 2 and 3 could change from one item to the next. Given that hints were contingent, the feedback that composed our hints was supposed to achieve the following goals:

1. Notifying the students that they chose the incorrect answer on their first try and urging them to listen and try the item again (mediation level 1).
2. Explaining the nature of the items, offering advice on the best strategies to address that type of question, and encouraging students to pay attention to the speakers' reactions to a change in sound or a particular response to a request or question (mediation level 2).
3. Providing learners with a transcription of a specific sentence from an audio file and asking them to choose the best option based on how they perceive the sentence in the context of the audio file (mediation level 3).
4. Presenting the appropriate response and elaborating on it (mediation level 4).

Each audio file had to be listened to twice before the learner could choose from the available options. If they did not correctly respond to a test item on their first attempt, the first hint would appear on the screen, instructing them to try again. If the students' second attempt was successful, they would receive three points and move on to the following question. If not, the second hint was displayed on the computer. The process

1. You hear a conversation about reading. The man enjoys reading books which

   A have characters that remind him of people he knows.

   B describe situations that he finds highly amusing.

   C narrate a wide range of fun stories.

   D are set in places that he is unlikely ever to visit.

   E describe serious and hardworking people all around the world.

Fig. 1 An item in the CDA test consists of a stem and five options

would continue if the students could not come up with the right response. After receiving each hint, a point would be subtracted from the overall score of each test item; thus, if a student received all four hints, they would receive a score of zero. An example of a listening question with its ZPD-based hints is shown in Fig. 2.

In our CDA project, an innovative extra step was taken to reduce the generated scores' guessing effect. In this regard, a guess-checking question that asked, "Why did you choose this option?" appeared immediately after an item had been properly answered on the first try. Then, two choices—one presenting the right explanation and the other the incorrect one—appeared on the screen. The program would move on to the next item if the test-taker selected the accurate description, assuming that the item had been completely mastered. Otherwise, it would be assumed that the test-taker had successfully responded to the item only by chance, which meant that special hints and prompts had to be given in the same way as they were to someone who had answered the question incorrectly. An example is illustrated in Fig. 3. Some screenshots of the website are also provided in the Additional file 1: Appendix.

Participants could receive their software-generated scoring profiles after finishing the test. The profile information is presented in Table 1.

In sum, learners' scoring profiles reported their scores and measured their success regarding the number of hints they used for each item in the test.

16. You hear a man talking about an activity holiday he went on as a child with his family. How did he feel during the holiday?

    A Pleased to be climbing with his father.
    B Upset with his father.
    C Disappointed with the rowing boat.
    D Curious about everything around him.
    E Bored by the climbing.

Hint 1. Your answer is wrong. Listen again to the conversation, please.

Hint 2. Your answer is still wrong. When answering **questions related to feelings of interlocutors**, you should look for sentences, phrases or vocabularies expressing the feelings you want to infer. You should pay attention to the speaker's tone of voice as well.

Hint 3. Your answer is still wrong. You can find the answer if you focus on the meaning of these sentences: <u>One day the weather stopped us going climbing. much to my relief.</u>

Hint 4. Your answer is wrong. The correct answer is "e". "<u>Much to my relief</u>" seems like an adverb phrase expressing the cheerfulness of a pleasant or happy discovery.

Fig. 2 A listening question, together with its ZPD-based hints

Guess-checking item (16): Why did you choose this option?

A) Because "his father was very keen on climbing" and this made him feel good. (incorrect)

B) Because "much to my relief" seems like an adverb phrase expressing the cheerfulness of a pleasant. (correct)

Fig. 3 A guess-checking item in the CDA test

Table 1 The scoring profiles of learners generated by the dedicated website

Score	Definition
1. The actual score	This score is almost the same as the conventional test scores and is calculated based on a pre-determined scale to make it comparable to the mediated test score (0 to 80 points, four points for each item).
2. The mediated score	This score is obtained by subtracting the score of hints each test taker uses from the overall hints' scores (80). It is worth mentioning that 1 point was subtracted for each hint.
3. The LPS (learning potential score)	It is calculated based on the formula developed by Kozulin and Garb (2002); LPS = (2*Mediated score _ Actual score) /Maximum score.
4. Hints used	The number of hints used for each item.
5. Missed items	These are items that each test taker might not have answered within the time allotted for each question.
6. Total time	The total time spent on the test.
7. The guessing effect score	It shows the total number of questions each learner has answered correctly by chance on the first try and their effects on the learner's actual and mediated scores.
8. The with-guessing-effect actual score	This score equals the actual score minus the guessing effect on the actual score (Four points are subtracted from the actual score due to recognizing any guesswork).
9. The with-guessing-effect mediated score	This score equals the mediated score minus the guessing effect on the mediated score (minus 1 point is subtracted from the mediated score due to recognizing any guesswork).
10. The with-guessing-effect LPS.	This score equals LPS minus the guessing effect on the LPS (the with-guessing-effect actual and mediated scores are replaced in the LPS formula to calculate this score).

2.   Semi-structured interviews

Due to the study's exploratory nature, interviews were conducted to collect data as they provide dynamic approaches through which participants can share their understanding of their world. Cohen et al. (2007) explained that the interview is not simply related to collecting data about life; it is a part of life itself.

The study's semi-structured interviews gave us an in-depth understanding of significant issues around the use of CDA in teaching and assessing inferential listening comprehension skills, as we were able to draw on the experiences of learners who have taken part in our online CDA test of inferential listening comprehension skills. The semi-structured interviews included questions about learners' perceptions of the usefulness of CDA in general and the CDA test employed in this study in particular.

Participants' personal preferences and experiences throughout the test and their suggestions for more effective CDA implementation were also investigated. Finally, the interviews ended by asking the participants about any concerns not addressed in the questions. The formulation of the interview questions is rooted in the research question and the literature review. Also, based on supervisor feedback and time estimates, the interview questions and protocol were adjusted until they were used in the interviews.

Data analysis

This study implemented a conventional approach to qualitative content analysis to derive the themes from the documented interviews. Thus, audiotaped interviews were transcribed, coded, and categorized into six major themes, each encapsulating one or more sub-themes, by going through the systematic approach of open, axial, and selective coding (Glaser, & G,, & Strauss, A., L., 1967). In other words, instead of imposing prescribed themes on the data, the researcher derived themes from the data collected (Kondracki et al., 2002). Hsieh and Shannon (2005) define qualitative content analysis as a research method for the subjective interpretation of text data content using a systematic categorization of coding and discovering themes and patterns.

The content analysis aims "to provide knowledge and understanding of the phenomenon under study" (Downe-Wamboldt, 1992, p. 314). The theme is defined by Boyatzis (1998, p. 4) as "a pattern found in the information that at a minimum describes and organizes the possible observations and at a maximum interprets aspects of the phenomenon." Conventional content analysis is typically employed with a research design whose goal is to describe a phenomenon, in this case, the test-takers' perceptions of the CDA test they have participated in. This data analysis approach is usually appropriate when the existing theory or research literature on a phenomenon is limited. Its advantage is that it allows researchers to obtain direct information from study participants without imposing pre-existing categories or theoretical perspectives (Hsieh & Shannon, 2005).

Since semi-structured interviews involved using open-ended questions (Hsieh & Shannon, 2005), the data analysis process began with a thorough review of all data to achieve immersion and gain a sense of the whole (Tesch, 1990). Next, the data were read word by word to extract codes (Miles & Huberman, 1994; Morgan, 1993; Morse & Field, 1995) and the exact words from the text that seemed to encapsulate key thoughts or concepts. Finally, the texts were explored again to check the whole analysis process. Inter-rater reliability was assured in this study by checking the consistency of the analyses between different researchers.

As the process progressed, labels for codes evolved that reflected more than one key concept. These were directly taken from the text and used as the basis for the first coding scheme. The codes were then categorized depending on how closely they were related and linked. The codes were organized and grouped into meaningful clusters using these emerging categories (Coffey & Atkinson, 1996; Patton, 2014). Based on the links between these larger subcategories, the researcher combined and organized them into smaller categories, later referred to as sub-themes and themes.

The researcher will reach a point when they must stop after going through the various stages of data analysis. According to Corbin and Strauss (2014), this saturation point is "the point when all the concepts are thoroughly defined and described," and no new information is gained. They claimed that the development of categories in terms of their attributes and dimensions is indicated by saturation. According to Ragin and Amoroso (2018), it is hard to identify this point in advance and they go on to say that researchers will only be able to do so once they are well-versed in the subject of their study. The analysis process continued until it had reached its saturation point where no more codes or categories could be extracted from the data analysis and no further analysis was necessary.

A tree diagram (Fig. 4) was then created to aid in organizing these categories into a hierarchical framework (Morse & Field, 1995). Exemplars for each theme and sub-theme were also discovered from the data and are presented in the results section.

Results

Content analysis of the gathered data resulted in the emergence of six major themes, each encapsulating one or more sub-themes. In what follows, the results are presented and explained.

As demonstrated in Table 2, the major themes included participants' perceptions of (1) the general features of the CDA test, (2) their personal experiences during the CDA test, (3) the effectiveness of the CDA test, (4) the brand-new feature of the CDA test, (5) the possible applications of the CDA test, and (6) their negative perceptions of the CDA test.

Discussion

The sub-themes for each category (reported in Table 2) are covered one by one in the following sections.

The test's appearance and its face validity

The first group of sub-themes, which captured participants' perceptions of the CDA test's appearance and its face validity, is presented here to exemplify learners' responses to the investigated items. Nearly all participants had a positive attitude toward learning through testing, describing it as a novel and engaging approach that strongly emphasizes understanding the material rather than getting a high grade. However, some students claimed that though learning through testing may be one of this method's strengths, it would cause stress and anxiety for those unable to answer the questions on their first few attempts correctly.

Additionally, as Poehner (2011) argues, the interventionist approach's scripted-based interactional behavior provides a chance for score comparison across participants. However, the scoring procedure does not necessitate a significant number of qualitative profiles. The quantification procedure supports the test's face validity, which makes it easier for examiners to satisfy both students and parents (Poehner, 2011). This further supports those learners who commented that the auto-generated scoring profiles have made them more eager to retake the test.
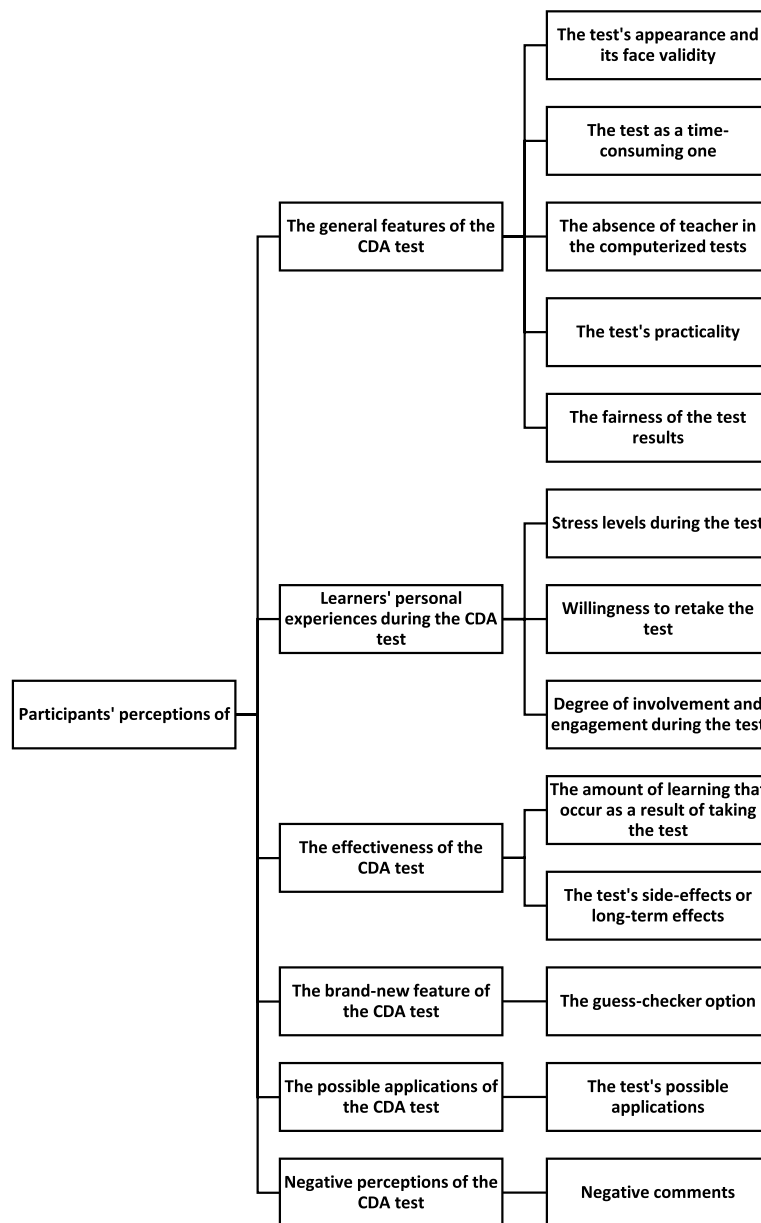
Fig. 4 Themes and sub-themes of the data gathered through the interview in a hierarchical framework

The test as a time-consuming one

Most participants agreed the test was a fantastic learning experience, even though it took longer than a typical test. Additionally, by lengthening the exam, a great opportunity was offered to review previously learned material that had been taught over a substantially longer period. The time, expense, and in-depth knowledge needed by teachers to use the DA technique in curriculum design, according to Haywood and Lidz (2007), may impede its widespread use. Despite the problem, Mehri and Amerian (2015) believe that if DA is utilized for educational reasons, a high-quality approach to training may reward the decision. Some participants stated that CDA assessments should not be viewed as time-consuming because extra time was only allowed for those who required

Table 2 Content analysis of the data gathered through the interview

Themes	Sub-themes	Excerpts
1. The general features of the CDA test	1.1. The test's appearance and its face validity:	• This exam made me feel like the main point is to learn the material, not to achieve a good mark. That was fantastic! • It was a novel and intriguing technique. A good attitude toward the exam is created through variety and novelty. • Depending on the individual's needs, hints were offered from indirect to more direct instructions. So, the difference between someone who simply needed a quick flip and someone who required a more thorough explanation was considered.
	1.2. The test as a time-consuming one:	• Although the test was time-consuming, the quality of the work produced was superior to that of a typical test. So, it seemed ok. • The exam time would be extended only for those who wanted further help, which was fantastic.
	1.3. The absence of a teacher in the computerized tests:	• The computerized test allowed me to get hints at my own pace. • Because the exam was computerized, human emotions were eliminated, the scores were not biased, and the error rate was minimized. • Computerization reduced the stress of being monitored by a teacher or an invigilator who might object to me.
	1.4. The test's practicality:	• It reduced the cost and time required to commute to school. • There were no limitations on how many times you could take the test or how many times you might participate in it.
	1.5. The fairness of the test results:	• In the classic exams, no point was lost for receiving hints, which was why the distinction between someone who needed help and someone who performed well independently was not included in the test results. • Because each question had five options plus a guess-checker, the degree of chance effect in the outcomes was substantially lower than in a traditional multiple-choice exam. • The error rate in announcing the results was meager because the human had no role in correcting it.

Table 2 (continued)

Themes	Sub-themes	Excerpts
2. Learners' personal experiences during the CDA test	2.1. Stress levels during the test:	• Because of the shift in our mindset toward the exam as a means of learning rather than an evaluation, the stress level was pretty low because of the comfort we acquired from the tips. • The fact that you do not answer any question suspiciously and do not pass it until you are sure it is answered correctly frees your mind and reduces stress.
	2.2. Willingness to retake the test:	• Knowing that the exam had an instructional aspect increased my desire to retake the test. • Knowing that we'd get our score right after the exam was over made us even more motivated to retake it. • The idea that I would not be penalized when I answered a question incorrectly and the hints would even assist me made me more eager to retake the exam.
	2.3. Degree of involvement and engagement during the test:	• This nature of the examination made me feel like I was playing a game; I was utterly immersed in the exam. • We were not distracted by the surroundings and were engaged in the test since we immediately received our performance results and realized our mistakes.
3. The effectiveness of the CDA test	3.1. The amount of learning that occurs as a result of taking the test:	• I learned the techniques and strategies needed to answer inferential questions with the help of hints. • While we were having fun with the hints, I believe some implicit learning occurred. • Hints eliminated the need to resort to additional resources to root out our problems, and we instantly identified the cause of our mistake.
	3.2. The test's side-effects or long-term effects:	• It internalizes this concept as the key is to learn the topic instead of getting a good grade in the long run. • In the long run, it teaches the students that there is a solution to every problem and that they must first master the required techniques to solve any problem. • Students become accustomed to not answering questions by chance after using guess checkers for a long time.
4. The brand-new feature of the CDA test	4.1. The guess-checker option:	• It was achieved successfully if the purpose was to ensure that the students had truly learned the information and taught the material via hints. • A guess-checker in the test indicated that the main goal was to think and understand the content, and the score was of secondary importance.

Table 2 (continued)

Themes	Sub-themes	Excerpts
5. The possible applications of the CDA test	5.1. The test's possible applications:	• We spent a long time learning something, and then with only a 1- or 2-h dynamic test, we could evaluate, review, recall, and establish a large amount of that content. • These types of tests, in my opinion, were ideal for instructional software, quizzes, midterm examinations, and any other form of exam which aimed to train and prepare students.
6. Negative perceptions of the CDA test	6.1. Negative comments:	• It would be preferable to be able to skip some questions and rearrange the order in which they are answered. • Because the computer could not interact emotionally with the participants, the sense of support provided by the teacher's presence was lost. • The fact that participants were not being monitored during the test raised the likelihood of cheating.

extra support. It is consistent with Mehri and Amerian's (Mehri & Amerian, 2015) claim that adopting CDA saves time and effort compared to earlier DA realizations.

The absence of a teacher in the computerized tests

Many students argued that the absence of a teacher from CDA exams was a good idea because it gave them the freedom to learn tips at their own pace, eliminated biases and error rates caused by human emotions, decreased the stress of being observed during the test and removed time, location, and accessibility barriers that a teacher might pose. Others argued that using CDA tests as a supplement to an accessible instructor rather than as the teacher's replacement would be preferable. These findings reinforce Tzuriel and Shamir's (Tzuriel & Shamir, 2002) comparison of the gains made by students who received both human and computerized mediation vs. those who only received human assistance. They found that learners who received both types of mediation benefited the most, which is not surprising.

Others supported what Sardrood and Javad (2011) revealed in their research on the Iranian EFL environment and said that using computers rather than teachers to deliver the mediation levels is a good idea since it eliminates the need for a qualified and experienced teacher. His research found that most teachers had a negative view of DA and thought implementing it in Iranian EFL classes would be difficult due to a lack of DA training, guidelines, technology resources, and large class sizes. However, as some of the participants in the current study have noted, using a computer to provide the mediation levels eliminates the need for lecturers to undergo significant training to carry out dynamic assessments fairly and transparently.

A participant said, "A computer can't analyze students' needs as effectively as a teacher, and it can't direct them precisely based on their needs. Besides, if I had a question, I could not get help from the computer." This statement echoes Mehri and Amerian (2015)'s claim that the prescribed standardized procedure for applying the

interventionist technique in CDA limits the mediator's ability to respond to the learners' needs. Harding et al. (2015) asserted that computers could only provide general and prescripted feedback; as a result, tailoring the hints to the learners' idiosyncrasies is sacrificed in favor of standardization, enabling CDA tests to be administered to an infinite number of students.

The test's practicality

Some students remarked that this approach is convenient because it requires less time than a conventional test conducted in a classroom setting. They all mentioned the test's time flexibility, noting that they could take and retake it whenever they were motivated or in the mood. Another advantage for some learners was the possibility of reviewing the hints and mediations. Others referred to the software's ability to simultaneously host as many test-takers as needed as a positive feature. Additionally, the related literature supports these advantages of CDA. For instance, Poehner (2008) claims that CDA provides the following advantages over DA: (1) it may be administered to many learners at once; (2) individuals can retake the test as frequently as they wish; (3) a report of each learner's performance is provided automatically.

The fairness of the test results

According to many participants, one of the most favourable aspects of the exam is the fairness of the CDA results. They claim that this fairness is achieved by providing the same hints to all test-takers gradually and based on their individual needs; reducing the same amount of score for receiving each hint; minimizing the guess effect on the scores by adding a guess-checker option to each question; eliminating stress, distraction, and environmental factors by giving test-takers a second chance to answer any question; and removing the effect of human biases on the scores. They added that these factors aided the test-takers in demonstrating their true abilities. Besides, the difference between someone who knew a question entirely, someone who knew it just partially, and someone who did not know it was well projected in the test scores.

   These comments align with Lantolf and Poehner (2004), stating that in dynamic assessment, fairness entails giving students feedback based on their scaffolding needs to reveal their full potential. They also support Nazari (2017)'s study on lecturer viewpoints on DA, in which some participants stated that DA would help their students more because it is fairer, less stressful, and produces a clearer picture of their abilities.

Stress levels during the test

While some learners cited certain test features as stressful, others mentioned different aspects, such as eliminating or removing stress. Therefore, some students suggested adding several features such as setting an available time for the entire exam rather than setting individually timed questions and providing an option to skip some questions or to rearrange the order of answering the questions, all of which could help them to get rid of their high-stress levels. Some of these comments are in support of, and others are in opposition to, Nazari (2017)'s study on lecturer perspectives about DA, in which some participants believed that DA is less stressful and produces a clearer picture of their abilities students' abilities. Furthermore, these suggestions can significantly help

future researchers improve their CDA software design to follow Malmir and Mazloom's (Malmir & Mazloom, 2021) claim that CDA tries to achieve maximum L2 learners' cooperation with test-makers by creating a non-threatening and learning-oriented setting.

Willingness to retake the test

Many participants cited certain aspects of the test as factors making them more eager to retake it. For instance, they referred to having an instructional aspect and going beyond the function of a typical test; being novel, nostalgic, technology-based, and somehow in a fun mode; not requiring a specific time and place with anyone to participate in the test; not penalizing the test-takers for answering a question incorrectly; providing them with a second chance; offering gradual hints while needed; producing fair and unbiased results; and offering the scores right after the exam was completed.

The findings of this part of the study were consistent with previous studies' findings (e.g., Adokh & Rafiee, 2017; Babamoradi et al., 2018; Dastjerdi & Taheri, 2016; Ebadi & Yari, 2015), showing that participants had positive attitudes toward the implementation of DA or CDA and found them effective in teaching or assessing English language skills. However, in this study, the learners' views were not entirely positive, and as one can see, there were several negative perceptions.

Degree of involvement and engagement during the test

Some students were pleased with their level of involvement in the test. They noted, for example, that the nostalgic and fun format of the test made them feel like they were playing a game, allowing them to become completely immersed in the exam. Others claimed that getting immediate feedback and instantly understanding their mistakes made the exam more challenging and helped them stay focused and avoid being distracted by their surroundings. However, several students reported not taking the test seriously since a teacher was not monitoring them.

The amount of learning that occurs as a result of taking the test

Regarding the design and the effectiveness of the hints, some stated they were helpful, informative, and instructive. Others noted that explanations for correct answers were very beneficial and informative because they helped them understand the reasons for their correctness. These comments further support Dunn and Lantolf (1998) notion that DA enhances the "learning to learn" criteria for instructional programs while improving the quality of learning. According to Mehri and Amerian (2015), one of the features of DA is metacognitive awareness, which is thought to make learners aware of their mental processes and hence improve their problem-solving skills.

Others stated that hints were efficient because they eliminated the need to ask the teachers or students questions whenever a mistake occurred. Others commented on the hints' nostalgic character, saying, "this type of mediation is nostalgic and reminds us of teachers' guidance in the classroom during our childhood tests, but the difference is that these hints are the same for all learners, so they are fairer than a typical guidance."

On the other hand, some learners thought the hints were ineffective or only served as a reminder or alert and did not accomplish anything else. According to Mehri Kamrood et al. (2019), the quality of ZPD-based mediation supplied in such tests is a significant

problem for all interventionist CDA procedures. In the piloting stage of the study, we tried to follow Poehner et al. (2014) in offering ZPD-based mediation by delivering the test in an interactionist format to some students so that we could have the best possible arrangement of hints through communicating with learners; however, we admit that this mediation mechanism is far from perfect.

The test's side effects or long-term effects

Regarding the side-effects or long-term effects of the test, they also listed some positive points, such as encouraging students and instructors to keep up to date with the most recent technological advances, boosting students' bravery and confidence in taking the exam because of the test's idea of always having a second chance after making any mistake; internalizing the concept in the long run because the key is to learn the content rather than getting a good grade; accustoming the students to not answering questions by chance as a result of long-term use of a guess-checker option; and some of them even mentioned less paper consumption.

   As there are always some far-sighted students in every class, assigning these long-term effects to the test can positively affect their and their parent's attitudes toward implementing CDA tests, which can improve their performance in the test, as proved by Dastjerdi and Taheri (2016). Also, some students mentioned that using these tests can lead to more independent learners in the long run, which is in line with Poehner (2007), stating that learners in DA are expected to take on greater responsibility and receive less help as they become progressively autonomous. However, some participants disagreed with this claim, noting that using CDA tests might encourage laziness and mediation dependency in the long run.

The guess-checker option

Considering the guess-checker option, certain participants acknowledged that it had worked since they had attempted to answer some questions by chance and the program detected it. Others mentioned that the degree of chance effect on the outcomes was far lower than in a regular multiple-choice exam because each question featured five alternatives plus a guess-checker. This brand-new feature of our CDA test could, in turn, enhance the test's fairness which is highlighted by numerous studies stating that teachers and test developers should pay close attention to test fairness and the elements that may jeopardize it (Abobaker et al., 2021; Ahmed & Ganapathy, 2021; Bachman, 1990; Burger, 2017; Wallace & Qin, 2021).

   In conclusion, in response to Mehri Kamrood et al. (2019)'s call for finding a solution to overcome the guessing effect as a common drawback of all the previously conducted CDA studies, the comments made by the students on the guess-checker item confirmed the findings of our project's first phase. They showed that the solution was successful in minimizing the impact of guessing on test scores.

The test's possible applications

Regarding the CDA tests' purpose and applicability, it should be noted that the literature has recognized two types of assessment purposes: summative and formative (Bachman, 1990). Practitioners argue that instruction and assessment practices should be

harmonized and integrated for formative assessment procedures to be more efficient in learning (Brandon & Quarin-Wright, 2011). While explaining the differences between DA and formative assessment (FA), Lantolf and Thorne (2006) point out that DA may be used as a formative process in the classroom. Many participants in this study saw CDA as highly relevant and perfect for formative assessments, quizzes, midterm examinations, and any other type of test intended to train and prepare students, which is in line with the claims mentioned by Poehner and Lantolf (2005).

Poehner et al. (2014) argue that learner profiles created at the end of CDA exams can provide instructors with detailed and immediately applicable information about the sorts of help that individual learners require for various constructs in the test. Our analysis revealed that some participants had a similar idea by arguing that "the exam results help you discover any conceivable form of support and the extent to which it is needed. Consequently, you may develop lesson plans for future training sessions depending on your students' exam results." These findings are also referred to in the definition of formative assessment by D'Anglejan et al. (1990), regarding it as the process of information gathering that will inform instructors and learners about the degree of success associated with their respective efforts in the classroom. Formative assessment enables teachers to assess students' strengths and weaknesses in connection to specific curricular objectives, guiding them in the organization and structuring of instructional materials.

Negative perceptions of the CDA test
The majority of participants perceived the CDA test as fun and amusing, novel and interesting, stress-relieving, nostalgic, learning-centred, and having a lot of long-term positive advantages, which could be applied to a wide range of subjects and situations. In contrast, a few others assumed it was time-consuming, stressful, and carelessly designed, resulting in limited learning, increasing the likelihood of cheating, encouraging laziness in the long run, and ignoring the benefits of social aspects of teaching. These comments further support Poehner and Leontjev (2018), stating that, despite the insightful results obtained from the CDA studies, CDA's efficacy needs further investigation.

Conclusion and implications
The importance of developing inference-making skills while teaching listening comprehension as the most challenging skill for L2 learners cannot be overstated. And CDA, as a new strand of DA, has been used more and more in listening comprehension research. It is also important to highlight that, in L2 research, guessing is considered a key problem when employing multiple-choice questions in CDA studies developed so far.

This study (which is a part of a more extensive CDA project delivered through a dedicated website, www.lingeli.com , focusing on inferential listening skills while trying to minimize the guessing effect) adopted a multiple case study design and undertook semi-structured interviews to investigate learners' perspectives on an online computerized dynamic assessment of inferential listening comprehension skills.

The study's results indicated that, in sum, positive perceptions outweighed negative ones, as most of them were mentioned by more than 75% of test-takers. However, the negative comments were beneficial and eye-opening as they contained suggestions that may help the researchers add or adjust some features and move forward in improving

CDA tests' design which is a necessity according to many of the previously done CDA studies, including Mehri Kamrood et al. (2019), and Poehner and Lantolf (2013).

Additionally, the study's findings can strengthen the validity of CDA projects by including learners' perceptions and opinions of dynamic assessment. They can also help researchers make this assessment approach more applicable depending on the viewpoints of its potential recipients. During the interviews, the research participants mentioned some challenges to fully implementing the CDA test's potential capabilities to improve their listening comprehension.

The findings of this study have some implications for future practice. First, most participants strongly desired to retake the CDA test and reported a favorable opinion of it as opposed to the traditional quizzes or mid-term exams. Therefore, universities and institutions need to rearrange teaching and testing hours by offering less teaching and more assessment-as-learning employing computerized dynamic assessment models like the one presented in this study.

Second, the study's results showed how CDA generates a learning environment where objectives for both teaching and assessment are synchronized and linked. Additionally, language teachers can create a learning environment seen as amusing, nostalgic, novel, and stress-free by implementing consecutive CDA examinations in a few sessions over each term. It could thus have a favorable wash-back effect during exam times, possibly resulting in more efficient learning. It implies that those who create curricula and course materials can create syllabi that support CDA as a teaching strategy. It is hoped that it may stimulate modifications within the discipline or institution (Norton et al., 2006). To achieve this goal, academic professionals need to take courses in which examples of CDA are presented. Because as Guskey (2000) notes, meaningful education improvements hardly ever occur without professional advancement.

Third, curriculums that apply the CDA approach can help language learners by offering mediation suited to their requirements, enabling them to uncover and realize their full potential and develop autonomy within the classroom and in future activities. According to the participants, it may also have some long-term benefits, such as keeping teachers and students up to date and boosting students' confidence in taking the exams. The CDA software also internalizes that learning the material is more important than getting a good grade. Decreasing the chance effect contributes to forming the habit of not answering questions by chance.

While the present study's findings may give a clear picture of the Iranian EFL learners' perceptions of the online computerized dynamic assessment of inferential listening comprehension skills, they should be interpreted with care due to the study limitations. For instance, this study did not examine potential gender differences or similarities in learners' perceptions. Different genders may view CDA and its possible applications from different aspects. Thus, another subject for future research could be investigating potential gender differences and similarities in English language learners' perceptions of CDA and its potential application in English language courses. Furthermore, in light of some learners' negative perception of the test's graphical appearance, additional work is required to improve the test and make it more appealing to learners, such as using avatars to offer audio hints during mediation. Finally, more research is required to examine learners' perceptions of CDA software across different proficiency levels. It will assist researchers in gaining more insight into how applicable, preferable, and acceptable CDA is to students with different proficiency levels.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40468-023-00221-9.

<div style="border:1px solid; padding:8px">

Additional file 1.

</div>

References

Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension. In J. P. Lantolf, & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of secondlanguages*, (pp. 57–86). London: Equinox.

Ableeva, R. (2010). Dynamic assessment of listening comprehension in second language learning [Doctoral dissertation, University of Pennsylvania State]. Ableeva_PS_Thesis_FINAL. https://etda.libraries.psu.edu/files/final_submissions/5374

Abobaker, R. M., Elsayed Khalil, S., Mubarak Merghani, M., Mahadeen, A., Abdelraheem, E. G., & Hamdan-Mansour, A. M. (2021). E-learning success factors from the perspective of academic staff at nursing and education colleges during COVID-19 pandemic: A comparative study. Educational Sciences: Theory and Practice, 21(3), 1–10. doi: https://doi.org/10.12738/jestp.2021.3.001.

Adokh, H., & Rafiee, M. (2017). Investigating the insiders' perspectives about dynamic assessment process and practicality. *International Journal of Research Studies*, *6*(2), 41–47.

Ahmed, A. A. A., & Ganapathy, A. (2021). Creation of automated content with embedded artificial intelligence: a study on learning management system for educational entrepreneurship. *Academy of Entrepreneurship Journal*, *27*(3), 1–10.

Al-Dawoody Abdulaal, M. A., Ramadan Khalil, N., Heji Alenazi, M., & Robso Wodajo, M. (2022). Dynamic vs nondynamic assessments: Impacts on intermediate EFL learners' receptive skills. *Education Research International*, *2022*, 1–12. https://doi.org/10.1155/2022/5372929.

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), 465–483. https://doi.org/10.2307/328585.

Alnabhan, M. (2002). An empirical investigation of the effects of three methods of handling guessing and risk taking on the psychometric indices of a test. *Social Behavior and Personality: An International Journal*, *30*(7), 645–652. https://doi.org/10.2224/sbp.2002.30.7.645.

Babamoradi, P., Nasiri, M., & Mohammadi, E. (2018). Learners' attitudes toward using dynamic assessment in teaching and assessing IELTS writing task one. *International Journal of Language Testing*, *8*(1), 1–11.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford university press.

Baxter, P., & Jack, S. (2015). Qualitative case study methodology: study design and implementation for novice researchers. *The Qualitative Report*, *13*(4), 544–559. https://doi.org/10.46743/2160-3715/2008.1573.

Boyatzis, R. E. (1998). *Transforming qualitative information: thematic analysis and code development*. SAGE.

Brandon, J., & Quarin-Wright, M. (2011). Student assessment policy and practice in Alberta: An assessment for learning. *Leading Student Assessment*, *15*, 59–88.

Burger, R. (2017). Student perceptions of the fairness of grading procedures: A multilevel investigation of the role of the academic environment. *Higher Education*, *74*, 301–320. https://doi.org/10.1007/s10734-016-0049-1.

Burton, R. F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: Question selection and guessing of answers. Assessment &Amp; Evaluation. *Higher Education*, *26*(1), 41–50. https://doi.org/10.1080/02602930020022273.

Burton, R. F. (2004). Multiple choice and true/false tests: reliability measures and some implications of negative marking. *Assessment &Amp; Evaluation in Higher Education*, *29*(5), 585–595. https://doi.org/10.1080/02602930410001689153.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: complementary research strategies*. SAGE.

Cohen, L., Manion, D., & Morrison, K. (2007). *Research methods in Education*. Rutledge.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*, (4th ed.,). SAGE.

D'Anglejan, A., Harley, B., & Shapson, S. (1990). Student evaluation in a multidimensional core French curriculum. *Canadian Modern Language Review*, *47*(1), 106–124.

Darhower, M. A. (2014). Synchronous computer-mediated dynamic assessment : A case study of L2 Spanish past narration. *CALICO Journal*, *31*(2), 221–243. https://doi.org/10.11139/cj.31.2.221-243.

Dastjerdi, H., & Taheri, R. (2016). Impact of dynamic assessment on Iranian EFL learners' picture-cued writing. *International Journal of Foreign Language Teaching and research*, *4*(13).

DonYei, Z. (2007). *Research methods in applied linguistics*. Oxford university press.

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care For Women International*, *13*(3), 313–321.

Drever, E. (1995). *Using semi-structured interviews in small-scale research: A teacher's guide*. ERIC.

Dunn, W. E., & Lantolf, J. P. (1998). Vygotsky's zone of proximal development and Krashen's i + 1: Incommensurable constructs; incommensurable theories. *Language Learning*, *48*(3), 411–442.

Ebadi, S., & Bashir, S. (2020). An exploration into EFL learners' writing skills via mobile-based dynamic assessment. *Education and Information Technologies*, *26*(2), 1995–2016. https://doi.org/10.1007/s10639-020-10348-4.

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, *32*(5–6), 527–555. https://doi.org/10.1080/09588221.2018.1527362.

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.

Ebadi, S., & Saeedian, A. (2019). Exploring L2 learning potential through computerized dynamic assessment. *Teaching English Language*, *13*(2), 51–78.

Ebadi, S., & Vakili Latif, S. (2015). Dynamic assessment of EFL learners' listening comprehension via computerized concept mapping. *I-Manager's Journal of Educational Technology*, *12*(2), 29–41. https://doi.org/10.26634/jet.12.2.3613.

Ebadi, S., Weisi, H., Monkaresi, H., & Bahramlou, K. (2018). Exploring lexical inferencing as a vocabulary acquisition strategy through computerized dynamic assessment and static assessment. *Computer Assisted Language Learning*, *31*(7), 790–817. https://doi.org/10.1080/09588221.2018.1451344.

Ebadi, S., & Yari, V. (2015). Learners' perspective on using dynamic assessment procedures in vocabulary knowledge development. *English for Specific Purposes World*, *48*(16), 1–22.

Estaji, M., & Saeedian, A. (2020). Developing EFL learners' reading comprehension through computerized dynamic assessment. *Reading Psychology*, *41*(4), 347–368. https://doi.org/10.1080/02702711.2020.1768981.

Farangi, M. R., & Saadi, Z. K. (2017). Dynamic assessment or schema theory: The case of listening comprehension. *Cogent Education*, *4*(1), 1312078.

Glaser, B., & G, & Strauss, A., L. (1967). The discovery of grounded theory: Strategies for qualitative research. *Adline de Gruyter*, *17*(4), 364.

Guskey, T. R. (2000). *Evaluating professional development*. Corwin press.

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: elaborating on diagnostic principles. *Language Testing*, *32*(3), 317–336. https://doi.org/10.1177/02655 32214564505.

Haywood, H., & Lidz, C. (2007). *Dynamic assessment in practice: clinical and educational applications*. Cambridge University Press.

Hidri, S. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context. *Language Testing in Asia*, *4*(1), 1–19.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

Jaradat, D., & Sawaged, S. (1986). The subset selection technique for multiple-choice tests: An empirical inquiry. *Journal of Educational Measurement*, *23*(4), 369–376. https://doi.org/10.1111/j.1745-3984.1986.tb00256.x.

Kao, Y. T., & Kuo, H. C. (2021). Diagnosing L2 English learners' listening difficulties and learning needs through computerized dynamic assessment. *Interactive Learning Environments*, *1–25*. https://doi.org/10.1080/10494820.2021.1876738.

Karandikar, R. L. (2010). On multiple choice tests and negative marking. *Current Science*, *99*(8), 1042–1045.

Karlström, P., & Lundin, E. (2013). CALL in the zone of proximal development: Novelty effects and teacher guidance. *Computer Assisted Language Learning*, *26*(5), 412–429. https://doi.org/10.1080/09588221.2012.663760.

Kern, R. (2000). *Literacy and language teaching*. Oxford University Press.

Kondracki, N. L., Wellman, N. S., & Amundson, D. R. (2002). Content analysis: Review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behavior*, *34*(4), 224–230.

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, *23*(1), 112–127. https://doi.org/10.1177/0143034302023001733.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, *1*(1), 49–72. https://doi.org/10.1558/japl.1.1.49.55872.

Lantolf, J. P., & Poehner, M. E. (2013). The unfairness of equal treatment: Objectivity in L2 testing and dynamic assessment. *Educational Research and Evaluation*, *19*(2-3), 141–157. https://doi.org/10.1080/13803611.2013.767616.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and genesis of second language development*. Oxford University Press.

Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education–Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, *39*(3), 188–193. https://doi.org/10.1016/j.stueduc.2013.07.001.

Li, M., & Zhu, W. (2013). Patterns of computer-mediated interaction in small writing groups using wikis. *Computer Assisted Language Learning*, *26*(1), 61–82.

Li, M., & Zhu, W. (2017). Explaining dynamic interactions in wiki-based collaborative writing. *Language Learning & Technology*, *21*(2), 96–120.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice.* John Wiley & Sons.

Ma, Q. (2017). A multi-case study of university students' language-learning experience mediated by mobile technologies: a socio-cultural perspective. *Computer Assisted Language Learning*, *30*(3-4), 183–203. https://doi.org/10.1080/09588 221.2017.1301957.

Mack, N., & Woodsong, C. (2005). *Qualitative research methods: A data collector's field guide*, (1st ed.,). Family Health International.

Malmir, A., & Mazloom, P. (2021). The impact of group dynamic assessment (GDA) vs. computerised dynamic assessment (C-DA) on Iranian EFL learners' pragmatic comprehension. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, *9*(1), 65–92.

Mameren, H. V., & Vleuten, C. V. D. (1999). The effect of a 'don't know' option on test scores: Number-right and formula scoring compared. *Medical Education*, *33*(4), 267–275. https://doi.org/10.1046/j.1365-2923.1999.00292.x.

Mashhadi Heidar, D., & Afghari, A. (2015). The effect of dynamic assessment in synchronous computer-mediated communication on Iranian EFL learners' listening comprehension ability at upper-intermediate level. *English Language Teaching*, *8*(4), 14–23. https://doi.org/10.5539/elt.v8n4p14.

Mehri, E., & Amerian, M. (2015). Challenges to dynamic assessment in second language learning. *Theory and Practice in Language Studies*, *5*(7), 1458.

Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2019). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, *34*(7), 868–897. https://doi.org/10. 1080/09588221.2019.1645181.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*, (An expanded sourcebook. ed.,). SAGE.

Modarresi, G., & Alavi, S. M. (2014). Examining the emotional influences of computerized dynamic assessment on EFL learners. *International Journal of Language Learning and Applied Linguistics World*, *7*(3), 138–155.

Morgan, D. L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research*, *3*(1), 112–121.

Morse, J. M., & Field, P. A. (1995). *Qualitative research methods for health professionals*. SAGE.

Nazari, A. (2017). Dynamic assessment in higher education English language classes: a lecturer perspective. *The Journal of Language Learning and Teaching*, *7*(1), 100–118.

Norton, L., Harrington, K., Norton, B., & Shannon, L. (2006). *Challenging traditional forms of assessment: University teachers' views on examinations. Making a Greater Difference: Connecting to Transformational Agendas*.

Oskoz, A. (2005). Students' dynamic assessment via online chat. *CALICO Journal*, *22*(3), 513–536 http://www.jstor.org/ stable/24147936.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE.

Pileh Roud, L. F., & Hidri, S. (2021). Toward a sociocultural approach to computerized dynamic assessment of the TOEFL iBT listening comprehension test. *Education and Information Technologies*, *26*(4), 4943–4968. https://doi.org/10.1007/ s10639-021-10498-z.

Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, *91*(3), 323–340.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer.

Poehner, M. E. (2011). Validity and interaction in the ZPD: Interpreting learner development through L2 dynamic assessment. *International Journal of Applied Linguistics*, *21*(2), 244–263.

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 (C-DA). *Language Teaching Research*, *17*(3), 323–342. https://doi.org/10.1177/1362168813 482935.

Poehner, M. E., & Leontjev, D. (2018). To correct or to cooperate: Mediational processes and L2 development. *Language Teaching Research*, *24*(3), 295–316. https://doi.org/10.1177/1362168818783212.

Poehner, M. E., Zhang, J., & Lu, X. (2014). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, *32*(3), 337–357. https://doi.org/10.1177/0265532214 560390.

Qin, T., & van Compernolle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, *25*(2), 55–74.

Ragin, C. C., & Amoroso, L. M. (2018). *Constructing social research: The unity and diversity of method*, (3rd. ed.,). SAGE.

Richards, K. (2009). Interviews. In J. Heigham, & R. A. Croker (Eds.), *Qualitative research in applied linguistics: A practical introduction*, (pp. 182–199). Springer.

Sardrood, E. H., & Javad, S. (2011). Dynamic Assessment in Iranian EFL Classrooms: A Post-method Enquiry. *Journal of English Language Pedagogy and Practice*, *4*(9), 47–63.

Shabani, K. (2014). Dynamic assessment of L2 listening comprehension in transcendence tasks. *Procedia-Social and Behavioral Sciences*, *98*, 1729–1737.

Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education*, *7*(2).

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: the nature and measurement of learning potential*. Cambridge university press.

Stuckey, H. (2013). Three types of interviews: Qualitative research methods in social health. *Journal of Social Health and Diabetes*, *1*(2), 56–59. https://doi.org/10.4103/2321-0656.115294.

Teo, A. (2012). Promoting EFL students' inferential reading skills through computerized dynamic assessment. *Language Learning & Technology*, *16*(3), 10–20.

Tesch, R. (1990). *Qualitative research: Analysis types and software protocols*. The Falmer Press.

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

Vakili, S., & Ebadi, S. (2019). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning*, *35*(3), 345–380. https://doi.org/10.1080/09588221.2019.1698616.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191.

Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge: Harvard University Press.

Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(1), 492–521.

Wang, P. (2015). The effect of dynamic assessment on the listening skills of lower-intermediate EFL learners in Chinese technical college: a pilot study. *Journal of Language Teaching and Research*, *6*(6), 1269–1279.

Yang, Y., & Qian, D. D. (2017). Assessing English reading comprehension by Chinese EFL learners in computerized dynamic assessment. *Language Testing in Asia*, *7*(1), 1–15. https://doi.org/10.1186/s40468-017-0042-3.

Yang, Y., & Qian, D. D. (2019). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, *33*(5–6), 628–652. https://doi.org/10.1080/09588221.2019.1585882.

Yin, R. K. (2003). *Case study research: Design and methods*, (3rd ed.,). SAGE.

Zimmerman, D. W., & Williams, R. H. (2003). A new look at the influence of guessing on the reliability of multiple-choice tests. *Applied Psychological Measurement*, *27*(5), 357–371. https://doi.org/10.1177/0146621603254799.

Publisher's Note