# Foreign Language Learning Assessment in the Age of ChatGPT: A Theoretical Account

**Marwan Saeed Saif Moqbel**
Ibb University, Yemen
marwan_s1977@yahoo.com
ORCID 0000-0001-6099-3997

**Abdu M. Talib Al-Kadi**
Ibb University, Yemen
findtalib@gmail.com
ORCID 0000-0003-3805-7507

## Abstract

In the Artificial Intelligence (AI) age now, foreign language learners can get unlimited support on their learning tasks from advanced AI chatbots, primarily the ChatGPT. At the same time, such a language chatbot intensifies the importance of rethinking learning outcome assessment. Traditional assessments that draw on a teach-and-test- approach are of little use and no longer valid for a comprehensive understanding of students' knowledge and skills that they, by and large, obtain in informal learning settings and contribute to their overall performance. Hence, alternative assessments can reflect more on learners' actual performance that likely involves AI chatbots and yet goes unnoticed by traditional assessment. This paper discusses alternative assessments regarding their nature, forms, characteristics, advantages, and integration into L2 programs. It establishes a foundational theoretical account for future research that would take place when ChatGPT becomes commonplace in the worldwide L2 contexts. It charts new research territories and passes the torch to second and foreign-language learning assessors to reflect on their teaching situations and reimagine L2 programs in light of the affordances of the ChatGPT, which has made a significant breakthrough in learning and teaching languages.

## Introduction

Teaching and learning are complex processes, taking place in various settings and different forms. These processes, which are rarely assessment-free, have been largely affected by technological inventions (Bravo et al., 2015). The most recent and vibrant technology is ChatGPT, which resulted from many attempts within the remarkable rise of Artificial Intelligence (AI), whose impact on education has become food for thought in academia (Ali et al., 2023; Fitria, 2023; Tuomi, 2018). According to Tuomi (2018), learning and teaching over the upcoming years are bound to change under the inevitable influence of AI. It enables new learning and teaching practices beyond the current teaching techniques, methods and approaches (Fitria, 2023; Hong, 2023; Rudolph et al., 2023; Tuomi, 2018; UNESCO, 2023). This inevitably requires teachers to engineer their teaching accordingly. The changes include AI-based teaching strategies and instructional materials that accommodate students' interests, styles, and needs (Tuomi, 2018), along with measurements of learning outcomes—learning that occur with or without teacher intervention or formal teaching.

ChatGPT, the most recent version of a natural-language system (Fitria, 2023), has quickly gained popularity owing to its ability to provide meaningful answers and detailed responses in various subject areas (Rudolph et al., 2023). Such an AI chatbot is shaping up as the hottest issue on the agenda of 2023. It can simulate human-like conversations (Fitria, 2023), and this increases the chances of its likeliness to be a valuable language-learning tool that provides personal tutoring, authentic conversations and interaction. It is viewed as a generator of language learning materials and models (Hong, 2023; Rudolph et al., 2023; UNESCO, 2023) with the potential to revolutionize the profession of L2 teaching (Kasneci et al., 2023). It can be beneficial for designing or updating curricula, lesson planning, and assessment (UNESCO, 2023; Weller, 2023). It can reduce the burden of manual exam design if invested wisely (Kasneci et al., 2023; Rudolph et al., 2023).

All that said, the effectiveness of AI applications (ChatGPT, a case in point) has raised concerns about language teaching and assessment, especially in contexts where students are typically assessed based on their learning product, chiefly through written assignments and exams. It raises concerns about authenticity, honesty, and plagiarism (Fei, 2022; Hong, 2023; Kasneci et al., 2023; Rudolph et al., 2023) because this chatbot can generate responses to questions and generate essays, etc. (Ali et al., 2023), which facilitates cheating (Volante et al., 2023). In this light, a reform of traditional assessments (Stobart, 2023) is necessary. It is important to pore over some alternative assessments (AAs) that could be invested in English language programs. This paper is a deep dive into AAs to provide insightful ideas for the refinement of learning assessment in the age of ChatGPT, which has already begun to shake traditional teaching and assessment to their foundations. It aims to put forward a theoretical account for future research that would take place when the new invention (ChatGPT) is well-recognized and appropriated in L2 contexts.

## Alternative Assessments

In the age of diversity of learning modes and models, evaluating L2 skills through traditional tests is hardly valid (Stobart, 2023). For this reason, teachers may want more valid and reliable tasks to identify what students can do in the target language. Given the recent technological advances–mainly the ChatGPT, teachers of foreign languages are now required to develop assessment tools that involve observable learning performance: making presentations or creating digital materials such as webpages, videos, and animations (Fei, 2022; Rudolph et al., 2023). The assessment should measure students' skills and knowledge in realistic, motivating, and authentic situations (Rudolph et al., 2023). Volante et al. (2023) suggested authentic assessments, including performance-based elements and observing students' learning across multiple contexts. For Hong (2023), L2 teachers can use tasks such as writing daily

journal entries. Such assessment forms are generally called alternative assessments (abbreviated in this paper as AA). It is an elastic term and thus should be decomposed for clarification (see Figure 1).
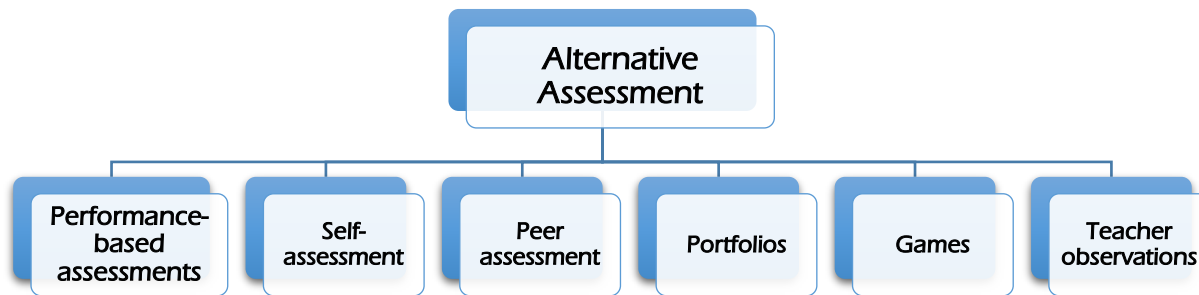


**Figure 1.** Forms of Alternative Assessment

## Performance-based assessments (PBA)

A common form of assessing learner outcomes is PBAs. A PBA is "an assessment activity that requires students to construct a response, create a product, or perform a demonstration" (McTighe & Ferrara, 1998, p. 34). It requires (a) a task to be performed or a product to be created and (b) some criteria for rating performances and products (Opp-Beckman & Klinghammer, 2006). The criteria should be well-articulated and shared with students through rubrics or scoring guides to enable teachers and students to monitor and profile students' language learning (Gottlieb, 2006). To Griffith and Lim (2012), sharing rubrics with students and communicating assessment standards increase their confidence and make them more engaged in learning.

PBA, integral to classroom teaching and learning, can allow students to express their learning directly and reflect real-life situations (Gottlieb, 2006). It creates opportunities for L2 students to produce authentic language, indicating what they know and can do in the target language (McTighe & Ferrara, 1998). To ensure the effectiveness of this type of assessment in L2 classrooms, teachers should use authentic tasks or activities directed with specific objectives and set some criteria for tasking students and evaluating their performance. Teachers should also consider students' interests and differences to ensure fair assessment and select appropriate tasks for students' levels and age. PBAs have several forms outlined in Figure 2: projects, role-playing, presentations, demonstrations, interviews, discussions/debates, writing samples, reports, the story of text retelling, cloze tests, and open-ended questions.

To begin with, a project is "an activity which focuses on completing an extended task or tasks on a specific topic" (Spratt et al., 2011, p. 33). Using projects as an assessment tool can enhance students learning as they accommodate various learning styles. A single student, a pair of students, a group, or an entire class may conduct projects. Using purposeful projects, teachers can meaningfully integrate the four language skills. Projects as assessment tools are used with young learners (e.g., designing posters or preparing pop-up books) and adult learners (e.g., making newspapers, bulletin boards, sketches, and news broadcasts). Using projects as an assessment tool enable teachers to identify what students can do with the target language. To be systematic, teachers have to use some assessment criteria with descriptions (e.g., appearance, quality, organization, the richness of ideas, etc.) and give them to students as a guide on how to progress in their projects and how their projects will be evaluated.
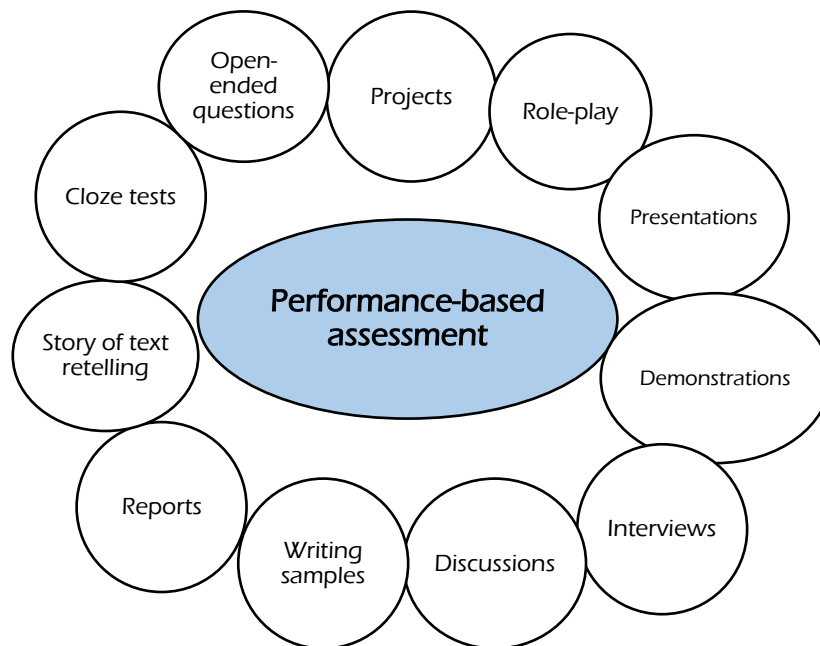
**Figure 2.** Forms of Performance-based Assessment

Role-play and simulations are another form of PBA. Role-play is a classroom activity in which students are given roles to act out in a particular situation, either in pairs or groups (Spratt et al., 2011). In contrast, with simulations, students act out real-life situations. These active learning activities can create interaction between students, allowing them to practice the target language in real-life situations (Encalada, 2018), which can help students develop their communicative skills. Role-play/simulation provides learners with opportunities to practice the knowledge and skills acquired and to be assessed on the knowledge and skills acquired before applying them in real-life settings. Although role-play or simulations are used mainly to evaluate speaking skills in L2, such assessment tools can be used by teachers to evaluate students' vocabulary, grammar, fluency, pronunciation, confidence, and motivation in L2 classes (Encalada, 2018; Phongsirikul, 2018). Using these activities as assessment tools, teachers need to develop meaningful assessment criteria or rubrics to maintain consistency and to help students understand the objectives of the simulation or the role-play. According to Encalada (2018), rubrics enable teachers to identify precisely the aspects that need to be improved for participating in conversations. To use role-play/simulation activities as assessment tools, teachers also have to develop activities involving learners to interact naturally in real-life situations (Encalada, 2018).

Presentations and demonstrations are another form of assessment based on performance. A presentation is an activity where a student gives a talk to their class (Spratt et al., 2011). Presentations can be of three types: controlled, guided, and free (Al-Issa & Al-Qubtan, 2010), depending on the level of students and purpose of the presentation. The common purpose of presentations in L2 classrooms is to create opportunities for students to practice speaking, help teachers integrate language skills, and promote learner-centeredness (Al-Issa & Al-Qubtan, 2010). Whether teachers evaluate students' presentations themselves (which is not always possible due to the long time required) or use peer assessment, the teachers may want to give students a clear idea about the assessment criteria. To better understand the requirements, a teacher may provide students with a copy of the rubric prepared in light of the assessment criteria and students' level (Moqbel, 2016; Phongsirikul, 2018). Demonstrations are additional

assessment forms with relevance to learners' performance. Demonstrations of learning in L2 classrooms can take the forms of writing (such as writing letters) or speaking (such as presentations). Such opportunities involving the target language demonstrate L2 learners' abilities to use particular words in contexts or specific grammatical structures in writing or speaking.

As interviews, discussions, and debates involve features of conversational interaction, they can be used to evaluate learners' ability to interact in authentic communicative situations (Fulcher & Davidson, 2007). Using interviews as assessment tools, teachers may invest in oral interviews to assess students' language proficiency and conversational communicative competence. For valid and reliable interviews, both structured and unstructured interviews should be directed by guidelines regarding topics and general questioning focus (Fulcher & Davidson, 2007). Interviews are an opportunity for elaboration, interruption, and abrupt change of topic (Hughes, 2003). ChatGPT creates opportunities to engage L2 students in virtual language conversations and interviews (Kasneci et al., 2023).

Discussions and debates are also fundamental PBA devices. While a discussion can be defined as a spoken interaction between three or four speakers who are then given a particular topic to discuss in the target language (Sybing, 2016), a debate is a formal method of interactive argument that often involves a moderator, audience, and the debate participants. Teachers can use discussions and debates to enhance students' speaking ability in L2. Teachers in L2 classrooms can also use these techniques to evaluate their students' verbal ability (fluency, pronunciation and vocabulary) and interactions. Concerning discussions as an assessment technique, a two-student discussion allows the teacher to assess students' performance more precisely. Discussions are usually centered on particular topics (Sybing, 2016). Yet, students may go beyond the topic as a natural discussion development. To ensure the validity and reliability of discussions/debates as an assessment technique, teachers should create appropriate scales for scoring that reflect the assessment criteria (Hughes, 2003). Teachers can use ChatGPT to create discussion prompts based on particular topics or events to engage their L2 students in meaningful discussions with ChatGPT (Kasneci et al., 2023).

Writing samples are also relevant to the assessment. Any piece of writing (creative writing, essays, reports, writing in response to prompts, a letter to an editor, etc.) that L2 learners produce can be used by their teachers to assess students' learning progress. The writing samples can include various topics, registers, and genres. According to their purposes, samples of writing can be scored either holistically; i.e., giving a single score based on an overall impression of the whole piece of writing or analytically, i.e., giving a separate score for each aspect or dimension of a task, such as grammar, vocabulary, etc. (McTighe & Ferrara, 1998). ChatGPT, producing written texts similar to human creation (Fitria, 2023), encourages students to depend on ChatGPT to do their writing tasks, which signals a cautionary note to teachers and educators. Teachers ask students to perform in-class pen-and-paper writing tasks to prevent learners' blind dependence on ChatGPT. However, it would be better if teachers could develop authentic writing tasks that encourage students to let their voices come out and make writing tasks more relevant to the students.

Ancillary to writing samples, reports can be of great significance when assessing L2 learners' performance. Reports, oral or written, are PBA activities which can be invested in assessing proficiency in L2. This is particularly useful for assessing students' speaking skills (Marzuki, 2017) as well as reading skills. For example, teachers can encourage students to read books and write simple book reports. Students may be tasked to report on particular events and present their reports orally in front of the whole class. The report writing technique can also be used to assess writing skills. Besides, the reporting technique helps assess the content knowledge. Whatever the purpose, teachers should develop a scale to assess students' oral or written performance. A checklist can also be used to evaluate how students report the topic (Marzuki, 2017).

Story-of-text retelling is an old and significant assessment tool. A story of text retelling is a post-reading/listening recalls in which readers/listeners tell what they remember orally/in writing (Morrow, 1989). Reading stories and then retelling them orally or in writing involves students reconstructing the text and making possible interaction among them, as Morrow (1989) stated. According to Praneetponkrang and Phaiboonnugulkij (2014), this technique is usually used in L2 classes for various purposes: oral retelling improves learners' comprehension and speaking skills and builds up their vocabulary while retelling in writing boosts learners' comprehension and writing skills and develop their vocabulary. The authors pointed out that the techniques used to involve students in story retelling include brainstorming, role play, and using pictures. Teachers can use this technique as an assessment tool in L2 classrooms to assess their students speaking and writing skills, reading and listening comprehension, and vocabulary.

What is more, a cloze test, as Fulcher and Davidson (2007) claimed, is a valid measurement of general language proficiency. It is a task type in which students read a text with missing words and try to replace the original words. The missing words are removed from the text at regular intervals, e.g., every seventh word (Hughes, 2003; Spratt et al., 2011). To complete the gaps correctly, cloze test takers are required to complete the gaps by replacing the original words, taking into account meaning and structure to find the answer. In predicting the missing words, cloze test takers must use the abilities that underlay all their language performance (Hughes, 2003).

Relevant to cloze tests, open-ended questions are a common PBA tool. An open-ended question is "a task or question that does not have a right or wrong answer but allows learners to offer their own opinions and ideas or to respond creatively" (Spratt et al., 2011, p. 29). Because open-ended questions launch conversations that allow students to hear and use language meaningfully, Wasik and Hindman (2013) called open-ended questions open-ended prompts. In L2 classrooms, the importance of using open-ended questions emerges from their active role in creating opportunities for students to use the target language meaningfully (Wasik & Hindman, 2013). To Wasik and Hindman, focused open-ended questions encourage students to use the vocabulary and ideas in a lesson. Besides, teachers can develop students' grammatical competence by encouraging students to use complete sentences; responses to open-ended questions require students to use more than one language skill and aspect. Hence, teachers can use such questions to assess different aspects, including speaking and reading comprehension, writing skills, vocabulary, grammatical competence, and knowledge.

## Self-assessment (SA)

With the emerging trends of learner-centeredness, learners can promote a sense of self-assessment— a process in which students actively evaluate themselves. This is believed to help learners better understand their learning outcomes and reflect on their performance, learning abilities and progress. Advantages of SA in L2 programs include improving language skills: writing (Meihami & Varmaghani, 2013), listening and speaking (Shahrakipour, 2014), and reading. It also encourages students' active participation in their learning and evaluation. Phongsirikul, 2018; Shahrakipour (2014) contend that self-driven assessment increases learners' motivation for learning and promotes their autonomy, independence, and lifelong learning skills. Students can be engaged in SA of their language skills through a variety of ways and techniques:  checklists, rubrics, reflection pieces (learning logs, journals, and dairies), conferences and interviews, self-correction (Wragg, 2004), progress cards, and computer-assisted assessment (Oscarson, 1989) (see Figure 2).
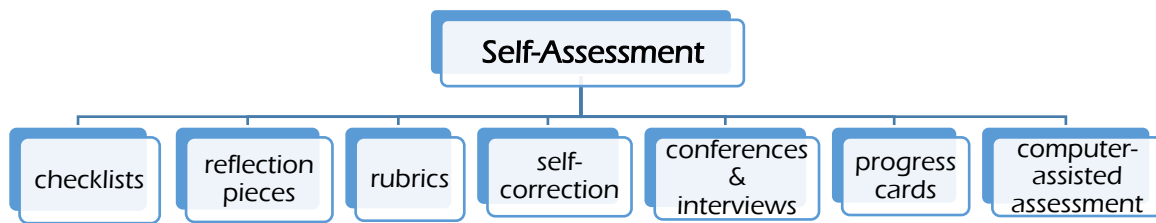
**Figure 3.** Techniques of Self-Assessment

To begin with, checklists are an example of SA tools. A checklist is "a list of dimensions, characteristics, or behaviours essentially scored as 'yes-no' ratings." It indicates that either a particular characteristic or behaviour is present or absent. Checklists often contain more dimensions to score than rating scales (Herman et al., 1992, p. 64). They can be "generic and applied to a skill or tailored to specific assignments" (NCDPI, 1999, p. 103). The problem with checklists is that no information is provided on the quality of student performance. The student ticks the level of performance that he thinks is appropriate according to his estimate (Oscarson, 1989). According to Wragg (2004), checklists primarily stimulate students' active learning through thinking about their answers and work, which may help them learn more effectively.

Reflection pieces are one more example of SA. They are "written entries in which students reflect on what they have learned and how they have learned it." They are valuable tools for students of all levels to monitor their learning and progress (NCDPI, 1999, p. 104). Reflection pieces like learning logs, journals, and diaries can take different forms. A learning log is a written account in which a learner can track his activities and progress through the term (Opp-Beckman & Klinghammer, 2006). In learning logs, students can reflect on what they have learnt, what they still have questions about, what was easy or difficult for them, what they must do to improve themselves, etc. Such reflection can be at the level of a particular concept or a unit of study. Journals are "daily or weekly writing entries by learners in which they reflect on their own learning experiences and progress" (Opp-Beckman & Klinghammer, 2006, p. 104). They "involve some reflection but are not as personal as a diary" (Falchikov, 2005, p. 11). Journals are usually written in narrative form. Using them as assessment tools can allow students to reflect on their learning, document their learning experiences, and express their feelings about it (Falchikov, 2005). A diary is a tool for reflection and analysis. It usually involves a written record (Falchikov, 2005) where students can reflect on their learning. According to Falchikov (2005), diaries are appropriate forms of assessment where the focus is on the learning process rather than the outcome. Diaries promote autonomous learning, encouraging students to take responsibility for their learning. Students can articulate their problems with course content through diaries.

Rubrics are tools including some criteria for the desired learning outcomes. They describe the performance standards linked to scales for grading students' performance. Typically, a rubric contains some assessment criteria and descriptors describing the proficiency or knowledge levels required for each criterion (Griffith & Lim, 2012). There are two main kinds of rubrics: holistic rubrics and analytic rubrics. The holistic rubric "evaluates the overall performance qualitatively."  It can provide one rating for a project or a performance and validate it with various criteria. Scores on such a scale give an overall impression of student ability or performance using a 3-, 4-, or 5-point scale (Griffith & Lim, 2012, p. 6). The problem with holistic scoring is that it gives students little feedback to help them achieve better (NCDPI, 1999). The analytic rubric breaks down the performance into different levels or components and scores them individually (Griffith & Lim, 2012; NCDPI, 1999). Points are then calculated to derive a quantitative measure of performance. For example, for a speaking task, a rubric might include the dimensions of pronunciation, fluency, verbal communication, non-verbal communication,

and confidence. Analytic rubrics can provide feedback on different components. They make students more aware of their strengths and weaknesses because the categories are rated separately. ChatGPT, according to Weller (2023), may act as a language teacher, analyze texts and provide supportive feedback based on a rubric, but this depends on how skilful use on the part of teachers who intend to employ ChatGPT in their assessment.

Self-correction is also a valuable SA tool. In self-correction, students are given an answer sheet, i.e., a key. However, this technique can only work when there is clarity and a single correct answer. Self-correction is mainly used in tests with closed answers, such as true/false or multiple-choice, rather than open-ended items. The self-correction method can be useless to students unless the teacher discusses the incorrect answers (Wragg, 2004).

Conferences and interviews are also relevant to SA. Conferences and interviews occur between students and the teacher: one-on-one, with several students, or with the entire class. Students can assess or talk about their learning using previously set criteria. They can even determine goals and expectations with their teacher (Opp-Beckman & Klinghammer, 2006). Besides, teachers can guide their students by asking questions, as NCDPI (1999) indicated.

Progress cards are an additional SA tool. The progress card is "a simple self-assessment tool used in many different educational settings" (Oscarson, 1989, p. 5). The progress cards are built on course objectives so that each group of objectives is graded to represent a difficulty level. Students proceed according to their improvement in a particular skill or learning area. Here, a student may use a personal test card on which he ticks off in the student column each language activity that he can perform, and the teacher ticks in the teacher column once he feels that the student can prove that (Oscarson, 1989).

SA can be computer-assisted. With the help of technology, many programs, materials, and applications have been developed for self-assessed language learning in different areas, such as vocabulary and grammar. These programs or applications provide tools for measuring students' mastery of the language. Usually, a learning hierarchy is formulated, and a diagnostic mechanism is built into such programs or applications so that either the learner himself or the program can decide when a review is needed. Such programs or applications may also have a testing feature which produces statements indicating each student's level (Oscarson, 1989).

These SA techniques give students more control over the learning process and increase their responsibility for learning (Gottlieb, 2006; Shahrakipour, 2014), self-confidence and make them more involved in learning (Gottlieb, 2006; Phongsirikul, 2018). More importantly, AS makes learners' more aware of the assessment criteria to (a) identify their weaknesses and strengths (Shahrakipour, 2014) and (b) monitor their improvement in areas of weaknesses. As in many other assessment tools, SA requires clear criteria when learners assess their performance or products to ensure unbiased assessment. Even after the implementation of SA, there should be systematic follow-up feedback from teachers on student's work (Meihami & Varmaghani, 2013).

## Peer assessment (PA)

In addition to self-assessment and performance-based assessment, peer assessment (PA) is an invaluable assessment form. Opp-Beckman and Klinghammer (2006) defined PA as an arrangement for students to "evaluate each other's work, using pre-set guidelines" (p. 104). The basic idea behind this assessment is to provide opportunities for students to evaluate each other's work more critically. When implementing PA, teachers may provide students with checklists, rating scales, or rubrics. For a successful implementation of PA, students should be trained in giving and using feedback and provided with a clear idea about PA. The teacher should give students clearly defined guidelines to assess each other's work and enough time for preparation, discussing and setting the assessment criteria clearly and making decisions about the PA techniques, i.e., a form, a checklist, etc. The teacher may even give the students a rubric

to understand the requirements better. After carrying out the PA, the teacher should give students constructive feedback on their performance.

Peng (2009) illuminated the benefits of integrating PA into L2 programs. It promotes students' autonomy and independence, increases their motivation, enhances their responsibility, and improves their self-confidence, as Phongsirikul (2018) endorsed. Peer assessment also helps students to develop collaborative skills, create opportunities for learning from each other, talk in the target language and get more feedback on their performance. With all these, PA can be essential in making students more aware of the course's objectives. According to Gottlieb (2006), PA promotes student involvement in their learning and helps them monitor their progress. Gottlieb added that PA is an effective means for having students practice the language with each other, which is vital in developing the target language.

## Portfolios

Another assessment device is portfolios. A portfolio, an authentic and practical assessment tool, can be used as a student learning assessment tool throughout a study or program (Lotfi, 2012; Narayan, 2023). It is, in Opp-Beckman and Klinghammer's (2006) words, "a collection of student work over a period of time" (p. 106). Essentially, it is a purposeful collection that gives teachers a clear idea about students' achievements, skills, abilities, and progress over a while and in one or more learning areas (Brown & Hudson, 1998). In portfolios as an assessment tool, the purpose and criteria of assessment should be defined, and what to put into the portfolio should also be determined (Herman et al., 1992; Narayan, 2023). The criteria can even be discussed with students to make students more aware of such criteria and more involved in the assessment process. Additionally, to ensure the effective use of portfolios, they should be developed from class tasks and connected to the course and its objectives. In portfolios, students can include whatever they believe to be essential for their learning process. They can include information, samples of work, and evaluations that serve as indicators of their performance, samples of written work (written stories, essays, etc.), tapes of oral work (role-playing, presentations, sketches, etc.), and checklists of tasks and performance (Lotfi, 2012; Narayan, 2023). Students can also include reflections on their works, such as learning logs or journals. Given the electronic age, portfolios can be in paper or electronic forms. In electronic portfolios, students use web 2.0 tools, such as wikis and blogs, to upload their works on the four language skills or their reflections on their performance.

Three main types of portfolios can be considered for classroom use: assessment portfolios, showcase portfolios, and collection portfolios. The first consists of items a student chooses to include according to specific assessment criteria (Lotfi, 2012). The items included in this type of portfolio are scored or evaluated. The second type contains students' best pieces or examples of work for each objective, and the third, also called a working folder, is a collection of all the pieces of a student's work during a particular time (Lotfi, 2012). In a relevant note, Rao (2006) noted that there is no single way to develop or implement portfolios. Rao proposed three portfolio categories representing a scheme for developing portfolios, namely collections, reflections, and assessment, stating that each category represents a distinctive stage of portfolio development and has criteria that reflect its function. Regarding collections, the decision about what to put in the portfolio is usually determined by the purpose of the portfolio (Burnaz, 2011).

According to Rao (2006), it is better to confine collections to one area, such as oral language development, reading competence, listening comprehension, etc. Reflections, which can take the form of a journal or diary, learning logs, self-assessment checklists, etc., can be on strategies of learning, students' attitudes and reactions (Rao, 2006), problem areas and difficulties, what has been learnt and what has still to be learnt, and improvement plans (Burnaz, 2011). The third category, i.e., assessment, is usually determined by the purpose of the portfolio, which should be defined clearly. If the purpose is to demonstrate growth or progress, the teacher can make judgments about the evidence of progress and provide those judgments

as feedback to students or make notes of them for their records. Similarly, students can use their portfolios to self-assess and monitor their progress. If the portfolio is to be used for an assessment, the teacher should determine when and how it should be evaluated and set clear assessment criteria (Rao, 2006). The reliability and validity of the contents should also be established and maintained. In addition, data collection should be systematic and in alignment with curriculum goals and objectives. Moreover, each piece of evidence should be linked to specified criteria as a rubric or a descriptive scale (Rao, 2006).

Using portfolios in L2 programs to assess students is valuable and practical. According to Lotfi (2012), portfolio assessment promotes students' autonomy, creates opportunities for students to take responsibility for their learning and plan their learning activities, and offers authentic information about their progress and performance, helping them monitor, reflect on, and reflect on their learning progress. In addition, portfolio assessment enables students to be involved in their learning and evaluation, promoting their motivation and helping them see their weaknesses, strengths, and development in different skills and areas (Burnaz, 2011). What is unique about portfolios is that they can be used in classes of any age or proficiency level (Opp-Beckman & Klinghammer, 2006) and can be used to assess all language skills and language systems.

## Games

Instructional games have become part of L2 teaching because they are a successful teaching strategy. Their importance is gained from the advantages they can provide in L2 classrooms. Instructional games can create opportunities for L2 learners to interact, communicate and practice the target language meaningfully and incidentally (Gozcu & Caganaga, 2016). They also create a stress-free learning environment, allowing students to learn in a relaxed and enjoyable atmosphere. Using games in the L2 classrooms purposefully, teachers can help their students learn the target language while engaging in the games and having fun (Bravo et al., 2015; Gozcu & Caganaga, 2016). Besides, using games in L2 classrooms creates an atmosphere of interaction and cooperation among students similar to that in the real world (Bravo et al., 2015), which can help L2 students learn the target language naturally.

Using games in the classroom should align the level of students of the target language and the stage of learning (teaching, reviewing, assessment). To be helpful, games should be used purposefully rather than for fun (Gozcu & Caganaga, 2016). Games are unique because they can be adapted to any situation, level and age, as Bravo et al. (2015) claimed. Using games as an assessment tool in L2 classrooms, teachers can evaluate their students' language skills and knowledge acquired during the lesson in a free-stress atmosphere and interesting and challenging manner rather than drilling. Making this process more concise and systematic requires establishing well-defined assessment criteria to be used by teachers and students as guidelines to assess students' learning and performance according to the established rubrics. Besides, the games used should be challenging and appropriate in terms of the level and age of students (Bravo et al., 2015).

## Teacher Observations

A formal assessment is not enough to make a reliable judgment about students' performance. Teacher observations in a stress-free condition in L2 classrooms are critical as they can be the base for reliable instructional decisions about students' performance (Ketabi & Ketabi, 2014). Teacher observations should be systematic; otherwise, they will result in useless information (Ketabi & Ketabi, 2014) and an unfair assessment of students' performance. To make observations more systematic, L2 teachers must plan their observations to decide what to observe, when, and how often to observe. Then, teachers have to think about how to record

their observations to ensure consistency of observations for all students. To record observations, teachers can use checklists, anecdotal records or rubrics (NCDPI, 1999).

Given the recent technological advances–mainly the ChatGPT, it is important to rethink L2 learning assessment, selecting from this long list of assessment tools that fit specific learning contexts. Alternative assessment widens the spectrum of evaluation and helps assessors get a comprehensive picture of students' overall language competence that reflect everyday learning situations (Opp-Beckman & Klinghammer, 2006; Moqbel & Al-kadi, 2020; Phongsirikul, 2018). AA is not limited to one method or technique but includes many forms and techniques ranging from simple tools (e.g., checklists) to long and complex ones that may take a semester or a year (e.g., portfolios). With this variation of tools, assessment should be part of the learning process rather than a predetermined summative assessment based on which success and failure in a program are traditionally determined.

## Implications

ChatGPT, construed as a cheat facilitator, heightens the importance of AAs. It provides new directions for language learners' competence assessment, representing their progress towards instructional goals and meaningful instructional activities. That said, there should be a shift from purely quantitative assessment to a mixture of qualitative and quantitative assessment focusing on students' essential content synthesis rather than assessing the learning product. To realise this, there should be a corresponding change in the teaching and learning process – a change that aligns with the digital technology assimilated to many daily activities that shape today's learners' lives. Given the tremendous potential of ChatGPT, this paper implies that teachers, besides familiarity with AAs, should be able to differentiate between Chat-GPT-generated and human-written products. L2 teachers can even use ChatGPT to generate assessment exercises for their students or measure their language competence while conversing with ChatGPT.

This implies a need for teachers with a ChatGPT mindset— teachers who devise state-of-the-art assessment tools instead of banning ChatGPT in their teaching or continuing with traditional assessment tool. They are supposed to be sensitized to the policies and the ethical use of ChatGPT and the consequences of academic misconduct. This also implies an overall shift in the teaching paradigm, which requires educators and education policymakers to revisit L2 pedagogy in their institutions and direct it toward more advanced modes that correspond well to the new pedagogies of projects implementing informal and personalized learning. Only when such shifts occur can we expect a change in the assessment procedures.

This brief account of assessment alternatives has implications for further Chat-GPT-based assessment topics. Each AA tool in this paper could be singled out for detailed exploration. This theoretical account might provide the impetus for more research on ChatGPT in terms of assessing second and foreign languages, attitudes and perceptions of teachers and students toward ChatGPT, and its impact on students' learning and teaching under certain conditions.

## Conclusion

AI applications in language learning and teaching have increased in parallel to new learning models and modes. Recently, ChatGPT has directed attention towards more reliable and valid assessment tools that gauge learning outcomes— humans' outcomes, not the machine-enabled performance generated by ChatGPT and other similar AI tools. The preliminary findings about ChatGPT in L2 education indicate that such a language chatbot is expected to invade academia and mislead educators when assessing learning outcomes. Besides teaching strategies involving ChatGPT, foreign language teachers may want to vary their assessment tools instead of the long-established testing approach. Given the affordances that ChatGPT has made accessible at low cost, teachers must reimagine and reform their

traditional assessment practices and adopt AAs, mostly practice-oriented assessments that minimize heavy reliance on ChatGPT. Teachers may employ ChatGPT that fed on big data, which has become a necessity, to collect information about their students' achievements and weaknesses, learning modes and strategies in such a way that helps them select proper AA tools. The entire learning and teaching process should be redesigned and re-assessed through AAs that bring in learners' formal and informal learning— learning occurs inside and outside the classroom and becomes inseparable from learners' overall abilities and performance. Educational policy-makers, teaching designers, and classroom/online teachers should realize the potential of AI for innovative assessment in correspondence to the tremendous technological advances that have already reshaped many life aspects, including L2 pedagogy.

## Disclosure Statement:

We (the authors of this paper) hereby declare that research ethics and citing principles have been considered in all the stages of this paper. Thus, we take full responsibility for the content of the paper in case of dispute.

## Ethics Statement

We confirm that the manuscript has been created by the authors and not an AI tool/Large Language Model (LLM). We are fully responsible for the content of the manuscript and are thus liable for any breach of publication ethics.

**Conflict of interest:**   We have no conflict of interest to declare.

**Funding:** None

## References

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41–49. https://doi.org/10.56540/jesaf.v2i1.51

Al-Issa, A. & Al-Qubtan, R. (2010). Taking the floor: Oral presentations in EFL classrooms. *TESOL Journal, 1*(2). 277-246. http://dx.doi.org/10.5054/tj.2010.220425

Bravo, I., Caicedo, E., & Soto, S. (2015). *Approaches to EFL teaching: Curriculum, culture, instruction, assessment, & technology*. Universidad Técnica de Machala.

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL QUARTERLY, 32*(4), 653-675. http://dx.doi.org/10.2307/3587999

Burnaz, Y. E. (2011). *Perceptions of EFL learners towards portfolios as a method of alternative assessment: A case study at a Turkish state university* [Unpublished master's thesis]. Middle East Technical University, Turkey.

Encalada, M. R. (2018). Role-plays as an assessment tool in English as a foreign language (EFL) class. In S. T. Soto, E. I. Palacios, & J. V. Holguín (Eds.), *Beyond paper-and-pencil tests: Good assessment practices for EFL classes* (pp. 49-73). UTMACH.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. RoutledgeFalmer.

Fei, L. V. L. (2022, December 16). *ChatGPT raises uncomfortable questions about teaching and classroom learning*. NIE National Institute of Education– Singapore. https://nie.edu.sg/docs/default-source/piar/media-coverage/20221216-chatgpt-raises-uncomfortable-questions.pdf?sfvrsn=cbb0657f_0

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *Journal of English Language Teaching, 12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book.* Routledge.

Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement.* Corwin Press.

Gozcu, E. & Caganaga, C. (2016). The importance of using games in EFL classroom. *Cypriot Journal of Educational Sciences, 11*(3), 126-135. http://dx.doi.org/10.18844/cjes.v11i3.625

Griffith, W. & Lim, H-Y. (2012). Performance-based assessment: Rubrics, web 2.0 tools and language competencies. *MEXTESOL Journal, 36*(1), 1-12.

Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment.* Association for Supervision and Curriculum Development.

Hong, W. C. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation, 3*(1). Preprint. https://jeti.thewsu.org/index.php/cieti/article/view/103/63

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 1-9. http://dx.doi.org/10.1016/j.lindif.2023.102274

Ketabi, S. & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory and Practice in Language Studies, 4*(2), 435-440. http://dx.doi.org/10.4304/tpls.4.2.435-440

Lotfi, H. (2012). Using portfolio assessment technique in EFL classrooms: Necessity, redundancy or luxury. *ELT Voices – India, 2*(6), 93-103.

Marzuki, A. G. (2017). Developing speaking skill through oral report in an EFL class in Indonesia. *Al-Ta Lim Journal, 24*(3), 243-254. http://dx.doi.org/10.15548/jt.v24i3.330

McTighe, J & Ferrara, S. (1998). *Assessing learning in the classroom.* National Education Association.

Meihami, H. & Varmaghani, Z. (2013). The implementation of self-assessment in EFL writing classroom: An experimental study. *International Letters of Social and Humanistic Sciences, 9*(2013), 39-48. http://dx.doi.org/10.18052/www.scipress.com/ILSHS.9.39

Moqbel, M. S. S. (2016). Alternative assessment in EFL classrooms: Why and how to implement it. In D. Boraie and A. Gebril (Eds.), *The proceedings of the 19th and 20th Nile TESOL/AUC Conference* (pp. 56-79). Cairo: Nile TESOL.

Moqbel, M. S. S. & Al-kadi, A. M. (2020). Alternative assessment in English departments at Yemeni universities: Attitudes and practices of faculty members. *Humanities and Educational Sciences Journal, 5*(12), 534-564. https://doi.org/10.55074/hesj.v5i12.182

Morrow, L. M. (1989). Using story retelling to develop comprehension. In K. D. Muth (Ed.), *Children's comprehension of text: Research into practice* (pp.37-58). International Reading Association.

Narayan, R. (2023). A critical review of portfolio assessment as an alternative tool in English language teaching classrooms. *English Language and Literature Studies, 13*(1), 1-8. http://dx.doi.org/10.5539/ells.v13n1p1

Opp-Beckman, L. & Klinghammer, S. J. (2006). *Shaping the way we teach English: Successful practices around the world (Instructor's manual).* University of Oregon.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(1), 1–13. http://dx.doi.org/10.1177/026553228900600103

Peng, J-C. (2009). Peer assessment of oral presentation in an EFL context (UMI Number: 3380148) [Doctoral dissertation, Indiana University]. ProQuest Dissertations and Theses database.

Phongsirikul, M. (2018). Traditional and alternative assessments in ELT: Students' and teachers' perceptions. *rEFLections*, *25*(1), 61-84. ERIC Number: EJ1271163

Praneetponkrang, S. & Phaiboonnugulkij, M. (2014). The use of retelling stories technique in developing English speaking ability of grade 9 students. *Advances in Language and Literary Studies*, *5*(5):141-154. http://dx.doi.org/10.7575/aiac.alls.v.5n.5p.141

Rao, Z. (2006). Helping Chinese EFL students develop learner autonomy through portfolios. *Reflections on English Language Teaching*, *5*(2), 113-122.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1), 1-22. https://doi.org/10.37074/jalt.2023.6.1.9

Shahrakipour, H. (2014). On the impact of self-assessment on EFL learners' receptive skills performance. *AJTLHE, 6*(1), 1-13.

Spratt, M., Pulverness, A., & William, M. (2011). *The teaching knowledge test (TKT) - glossary*. Cambridge University Press.

Stobart, G. (2023). Alternative assessment. In R. J Tierney; F. Rizvi & K. Ercikan (Eds.), *International Encyclopedia of Education* (4th ed.) (pp. 96-102). Elsevier.

Sybing, R. (2016). Structure for fostering discussion skills in the EFL classroom. *Journal of the Nanzan Academic Society*, *99*, 221-229.

Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education: Policies for the Future*. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository

UNESCO. (2023). *ChatGPT and Artificial Intelligence in higher education: Quick start guide*. UNESCO.

Volante, L., DeLuca, C., & Klinger, D. (2023, February 27). *ChatGPT and cheating: 5 ways to change how students are graded*. Queen's University. https://educ.queensu.ca/chatgpt-and-cheating-5-ways-change-how-students-are-graded

Wragg, E. C. (2004). *Assessment and learning in the secondary school*. Taylor & Francis e-Library.

Wasik, B. & Hindman, A. (2013). Realizing the promise of open-ended questions. *The Reading Teacher*, *67*(4), 302–311. http://dx.doi.org/10.1002/trtr.1218

Weller, D. (2023). ChatGPT for language teachers. David Weller.

**Auhtors:**

**Marwan Saeed Saif Moqbel** is an associate professor of Applied Linguistics at Ibb University, Yemen. He received a doctorate in English from SRTM University in 2015. He has about 20 years of experience in ELT. He has published eight research papers on various research areas, including teaching English and alternative assessment. He is interested in TESOL, L1 and L2, alternative assessment, technology-based teaching, gamification, and teacher education. He has participated in several local and international conferences. Besides, he serves as a reviewer for some scientific journals.

**Abdu Al-Kadi**, an assistant Prof. of Applied Linguistics, has taught English in multiple contexts over two decades. He has published on and continues to be agitated by issues within the realm of formal and informal language learning, TESOL, and post-method pedagogy. He has authored, co-authored, reviewed, and edited articles published in scholarly journals. He is a member of the TESOL International Association and the American Association for Applied Linguistics. He serves as a member of the International Editorial Board and reviewer of five scholarly journals.