Reconsidering Multiple-Choice Items for Speaking Assessment: Evidence from Libyan Secondary Schools

Ebraiek Deen Aljheme [1*], Mabrukah Abdullah Misbah [2]

[1,2] English Language Department, Faculty of Languages, Sabha University, Libya

إعادة النظر في أسئلة الاختيار من متعدد لتقييم التحدث: أدلة من المدارس الثانوية الليبية

ابريك دين حمد [1*]، مبروكة عبد الله مصباح [2]
[2،1] قسم اللغة الإنجليزية، كلية اللغات، جامعة سبها، ليبيا

[*]Corresponding author: ebr.aljheme@sebhau.edu.ly

Abstract

Speaking is a critical skill in English language learning. To be proficient speaker devoted time and effort in addition to practical use of language in real-life situations is required. In order to follow learners' improvement in speaking, a precise assessment tool of speaking skills must be utilized. However, in the context of Libyan secondary schools, speaking achievement tests rely on the use of multiple-choice (MCIs) formats as an assessment tool. MCIs have been a dominant format in testing due to their efficiency, reliability, and ease of scoring for testing different skills. Yet, when it comes to speaking, the validity of this format is criticized as means of evaluating actual speaking abilities. This study examined the validity of using multiple-choice items to assess speaking proficiency of Libyan secondary school students. A mixed method study was conducted. Three instruments were used for collecting the data: (analysis of speaking achievement test samples, a questionnaire, and an oral test administered to secondary school students). The findings of the study revealed that multiple-choice tests do not provide a valid assessment of Libyan secondary school speaking proficiency as they fail to capture the essential aspects of oral communication such as fluency, pronunciation, and interactive ability.

Keywords: Achievement Test, Multiple Choice Items (MCIs), Oral Proficiency, Speaking Assessment, Valid Test.

الملخص

التحدث مهارة أساسية في تعلم اللغة الإنجليزية. لكي تكون متحدثًا متقنًا، يلزم تخصيص الوقت والجهد بالإضافة إلى الاستخدام العملي للغة في سياقات الحياة الواقعية. ومن أجل تتبع تقدم المتعلمين في التحدث، يجب استخدام أداة تقييم دقيقة لمهارات التحدث. ومع ذلك، في سياق المدارس الثانوية الليبية، تعتمد اختبارات تحصيل التحدث النهائية على صيغ الاختيار من متعدد (MCIs). كانت هذه الصيغة (الاختيار من متعدد) صيغة سائدة في الاختبارات واسعة النطاق نظرًا لكفاءتها وموثوقيتها وسهولة تصحيحها لاختبار مهارات مختلفة. ومع ذلك، عندما يتعلق الأمر بالتحدث، فإن صحة هذا الشكل من الامتحانات موضع تساؤل في قدرته لتقييم قدرات التحدث الفعلية. تحققت هذه الدراسة من صحة استخدام عناصر الاختيار من متعدد لتقييم كفاءة التحدث لدى طلاب المدارس الثانوية الليبية. أجريت دراسة مختلطة. تم استخدام ثلاث أدوات لجمع البيانات: (تحليل عينات اختبار تحصيل التحدث، واستبيان، واختبار شفوي تم إجراؤه على طلاب المدارس الثانوية). وكشفت نتائج الدراسة أن اختبارات الاختيار من متعدد لا تقدم تقييما دقيقا لكفاءة التحدث في المرحلة الثانوية الليبية لأنها تفشل في قياس السمات الأساسية للتواصل الشفهي مثل الطلاقة والنطق والقدرة التفاعلية.

الكلمات المفتاحية: الاختبار التحصيلي، أسئلة الاختيار من متعدد، الاختبار التقييمي لمهارة المحادثة، اختبار فعال، الكفاءة الشفوية.

1. Introduction

No doubt that the English language is the most dominant global language, widely used across international communication, media, education, and digital platforms. Its influence continues to expand, with spoken English gaining particular significance due to the increasing demands of global communication and real-time interaction (Morozova, 2013). Speaking is a crucial component of language proficiency. It enables learners to express meaning and negotiate what they understand in real-life communication. Speaking is a key reflection of communicative competence and the ultimate goal in second language learners.

Chaney, (1988) and Gebhard, (1996) argue that speaking is one of the two productive skills in language learning alongside writing and is defined as the process of constructing and sharing meaning through verbal communication. Furthermore, Nunan, (2003) describes speaking as the production of systematic verbal utterances to convey meaning. Despite its fundamental role in language acquisition, speaking still one of the most difficult skills to be assessed accurately in English as a Foreign Language (EFL) context. As Knight (1992) points out, the complexity of assessing oral proficiency often leads teachers to rely on simplified formats, such as written tests, or avoid assessing speaking skills altogether.

Testing, according to Brown (2003), is a systematic procedure for measuring a learner's ability, knowledge, or performance in a given domain. In language learning, tests are used to measure learners' competence in the target language. One of the most commonly used test formats is the Multiple Choice Item (MCI). This kind of tests presents learners with a question and several possible answers, in which only one is correct. McNamara, (2000) states the distracters of MCIs typically represent common learner errors or misconceptions. While this format is convenient and easily scored by machines, it assesses recognition rather than actual language production, making it questioned for evaluating speaking competence.

Recent studies have reinforced the inadequacy of MCI formats for assessing speaking. For instance, Siwathaworn & Wudthayagorn, (2018) demonstrated that dynamic assessment (DA) in which the examiners interact with learners to check how much a learner can improve with help substantially enhanced the speaking performance of Thai university students, reinforcing the importance of roles in promoting communicative competence. Similarly, Maryam and Jalil (2020) revealed that DA positively affected Iranian EFL learners' speaking accuracy, although its effect on fluency was limited, highlighting the value of integrating interactive, formative assessments in the classroom.

In a more comprehensive review, Phung, Qin, & Nguyen (2020) called for a change from traditional test formats towards assessments that emphasize authenticity, real-life communicative demands, and sociocultural relevance. Furthermore, Al-Jamal and Al-Jamal (2023) questioned the validity of MCIs formats in assessing speaking in Jordan, it was revealed that although oral interviews and communicative tasks are preferred by many teachers, over 65% of them still depended on multiple the use of multiple choice formats to measure speaking competence highlighting a gap between pedagogical ideals and actual classroom practices. Additionally, Ghazali and Ahmad (2024) reported that student centered assessment tasks such as classroom seminars and peer evaluations encouraged greater learner autonomy and resulted in improved speaking proficiency in EFL settings. In the Libyan setting, Al-Badri (2018) reported that teachers depend widely on MCIs due to their ease of scoring, but oral skills are often neglected in the evaluation process. Al-Zadma Al-Fourganee (2018) argues teachers frequently prefer written tests to direct observation of speaking, citing practical constraints such as large class sizes and limited training, which prevents accurate assessment of students' communicative abilities. Hassan (2019) found that MCIs control achievement tests, in which language is tested in isolation, ignoring interactive and real-life communicative dimensions. All previous studies highlighted the inefficiency of using MCIs formats to assess speaking competence and the gap between pedagogical ideals and classroom practice. According to Brown (1994, 2003), valid speaking assessment must address multiple aspects such as fluency, pronunciation, rhythm, stress, intonation, and interactive capability. These aspects cannot be accurately assessed through multiple choice tests, which lack the capacity to assess spontaneous speech and authentic interaction. Building on this gap, the current study investigates the validity of MCIs formats for assessing speaking proficiency, examines oral proficiency through performance-based tasks, and explores teachers' perceptions of both traditional and alternative assessment methods in the Libyan context.

1.2 Research Questions

This study seeks to address the following primary research question:

Q1- To what extent are Multiple Choice Items (MCIs) formats a valid tool for assessing the speaking competence of Libyan secondary school students?

Q2- Which aspects of Libyan students' oral proficiency can be demonstrated when assessed through an oral proficiency test?

Q3- What are Libyan secondary school English teachers' perceptions of the effectiveness and limitations of using (MCIs) to assess speaking proficiency?

Problem statement

The use of Multiple-Choice Items in assessing speaking skills has long been criticized by many researchers. Brown (2003), Heaton (1995), and Nunan (1991) argue that this format of assessment is limited to testing specific linguistic features like grammar recognition. Crucially, MCIs formats fail to measure learners' communicative ability. Elabbar, (2011) argue that in many educational contexts, including Libya, assessment tends to prioritize written exams and objective formats. This is due to different constraints as the large class sizes, limited teacher experience and training, and the stress for standardized, easily scored tests. While MCIs may suggest a practical

solution, examiners may risk misrepresenting learners' true communicative ability. This is supported by Alderson (2000) who stresses; practicality should not outweigh validity in language testing.

Multiple Choice Items formats can provide indirect indication of knowledge that underlies speaking skill. MCIs cannot be considered as a valid standardized tool for measuring learners' speaking competence. Speaking achievement tests in Libyan secondary education system rely heavily on MCIs which risks reducing the construct of speaking competence to mere recognition of linguistic forms, ignoring crucial aspects of oral communication.

1.4 Scope of the Research
The research focuses specifically on the use of MCIs formats to test speaking competence of Libyan secondary school students. It questions the appropriateness of using this method and seeks to explore more effective alternatives for assessing speaking skills.

1.5 Research Hypothesis
The study hypothesize that speaking skill can be assessed more validly and reliably through other alternatives to MCIs such as performance-based tasks that may include face-to-face interviews, dialogue completions, and picture-cued speaking tasks. These interactive formats elicit direct language production, thereby providing a more reliable representation of learners' speaking abilities.

1.6 Aim of the study
This study aims to investigate the validity of using the Multiple Choice Items formats to assess speaking as a productive language skill. The study recommends alternative assessment tools that provide a more reliable and valid measurement of students' speaking competence in Libyan secondary school setting.

1.7 Significance of the Research
The significance of this study lies in two contributions: it questions existing assessment practices while proposing practical and pedagogically grounded alternatives to improve the assessment of speaking competence in the Libyan EFL secondary school context.

Literature review
2.1 Introduction
Luoma, (2004) argues that as a productive skill, speaking is widely recognized as one of the most challenging skills to assess in second language learning. Brown, (2004) claims that in contrast to the other receptive skills (reading or listening), speaking is productive, interactive, and multidimensional skill that involves fluency, accuracy, pronunciation, pragmatics, and discourse management. The complexity of measuring speaking proficiency raises questions about whether objective test formats such as Multiple Choice Items (MCIs) are valid to assess learners' speaking ability.

2.2 Defining Speaking
(Brown, 1994; Burns & Joyce, 1997) define speaking as an interactive process of constructing meaning that involves producing, receiving, and processing information simultaneously. Chaney (2006) further defines speaking as "the process of building and sharing meaning through the use of verbal and non-verbal symbols in a variety of contexts." This highlights the fact that is both dynamic and contextual.

Furthermore, Brown, (1995) argues that speaking is central to human daily interaction, However, acquiring effective speaking skills whether as a first or a second language requires a detailed understanding of when, why, and how to communicate. Effective speaking involves not only linguistic competence of the language as (grammar, pronunciation, and vocabulary) but also sociolinguistic competence of the target language, which concerns appropriate use of a language in relevant social contexts (Burns & Joyce, 1997; Carter & McCarthy, 1995; Cohen, 1996). A proficient speaker integrates all these competences to achieve communicative success.

Dostal (2007) argues that effective speaking starts with careful planning during which learners organize their ideas to express a coherent main message in speech production. This view is reinforced by Zhang and Lu (2021) who explain that strategic planning significantly enhances learners' fluency and coherence in speaking tasks.

2.3 Types of Speaking
Brown (2003) classifies speaking into five types based on task complexity and communicative purpose:
1- Imitative Speaking
This kind of speaking Involves repeating or parroting a word, phrase, or sentence for the sake of clarity and accuracy. Common classroom activities include drills (e.g., repeating "Excuse me" or "Can you help me?").
2- Intensive Speaking

It focuses on producing brief oral stretches that demonstrate control over specific phonological or grammatical features, such as intonation or stress. Examples may include sentence repetition, reading aloud, and dialogue completion.

 3- Responsive Speaking

This kind involves brief replies or responses to questions or prompts such as greetings and small talk. This type assesses comprehension and interaction at a basic level.

 4- Interactive Speaking

It is characterized by longer or more complex exchanges that include multiple turns or participants. It includes: Transactional exchanges for information transfer (e.g., interviews, role plays).
Interpersonal exchanges aimed at social relationships (e.g., casual conversations).

 5- Extensive Speaking

It consists of extended monologues tasks such as speeches, oral reports, or storytelling that needs planning and organization.

Recent research has extended these categories by emphasizing tasks highlighting the importance of authenticity and communicative functions. For example, Nguyen and Tran (2022) recommend integrating real-life tasks, such as problem-solving dialogues and peer discussions, to better train learners for authentic communication.

2.4 Assessing Speaking Competence

Luoma, (2004) and O'Sullivan, (2012) argue that authentic assessment methods for testing speaking proficiency typically include oral interviews, role plays, discussions, or integrated performance tasks. These tasks allow speaking examiners to evaluate fluency, interaction, and real time processing skills that MCIs cannot assess. Swain (1985) further argues, in her Output Hypothesis Theory, that the production of a language is essential to developing and demonstrating learners' communicative competence, and that is something MCIs formats fail to capture.

Effective speaking assessment requires clear criteria. Brown, (2003) proposes taxonomy of micro- and macro-skills that are essential for oral proficiency assessment:

1- "Micro-skills involved in speaking assessment include: producing English phonemes and allophonic variants, generating language chunks of different lengths, and employing appropriate stress, intonation, and rhythm, using reduced forms correctly, employing adequate vocabulary for specific communicative purposes. They also include maintaining fluency at varied speeds, using self-monitoring techniques and repair strategies, using correct grammar, producing natural speech elements, conveying meaning through diverse grammatical forms, and effectively employing cohesive devices."

2- "Macro-skills involved in speaking assessment include: fulfilling communicative functions suited to specific contexts, applying sociolinguistic features such as turn-taking, politeness, and style, linking ideas coherently, incorporate nonverbal cues including gestures and facial expressions, employing strategic devices like emphasis, rephrasing, and checking comprehension."

 Halliday, (1989), Burns & Joyce, (1997) and Solcova, (2011) all assert that, given these complex demands, speaking proficiency requires direct speech production rather than written responses. Halliday, (1989) further argues that spoken language is dynamic and typically less lexically dense than written language, placing more focus on actions and processes rather than nominalized forms. Despite developments, challenges in constructing reliable and valid speaking assessments exist. Hughes (1989) highlights enduring gap between communicative language teaching principles and practical oral testing. In addressing this, Canale (1984) proposed an Oral Proficiency Interview framework consisted of four stages:

1- Warm-up: involves small talk to reduce learners' anxiety.
2- Level check: involves narrative tasks like answering WH-questions, reading aloud, and guided role-plays.
3- Probe: involves more complex interactions including open-ended role plays and impromptu presentations.
4- Wind-down: involves feedback and questions.

 More recent studies have also emphasized the gap between the principles of Communicative Language Teaching (CLT) and the practical application of oral assessment methods. For instance, Fauzi and Ridwan (2025) found that while CLT principles significantly enhance students' speaking proficiency by promoting an interactive learning environment, challenges remain including preliminary reluctance to participate and the need for well-structured activities. In alignment with these pedagogical concerns, rapid advances in technology have introduced new possibilities for assessing speaking. More recent research has focused on enhancing test validity and learner engagement through dynamic and technology supported assessments. For instance, Li et al. (2023) found that using digital speaking portfolios in proficiency testing enhanced assessment reliability and student motivation. Furthermore, adaptive speaking tests supported by artificial intelligence (Xu et al., 2025) present promising developments for personalized and scalable oral assessment. Ma et al. (2025) demonstrated that speech-based large language models outperform earlier automated scoring systems by providing stronger generalization across diverse tasks. Building on this, Fang et al. (2025) proposed multimodal large language models (MLLMs) that integrate audio and text through a curriculum-based, speech-first training approach. Their model gained

significant improvements in holistic oral assessment, particularly in delivery-related dimensions. Similarly, Wang et al. (2023) strengthened assessment reliability by embedding measures of content relevance and detailed grammar error detection into hybrid scoring models.

Beyond mechanical scoring, recent innovations have also explored learner-oriented applications of large language models. Cha et al. (2024), for instance, developed CHOP, a ChatGPT-based interactive platform. This application is designed for oral presentation practice. The system provides personalized feedback to enhance speaking development while simultaneously reducing the anxiety often and it is often associated with oral exams. These recent developments highlight the need for a reevaluation of oral assessment practices to align more closely with communicative competence and to incorporate technological advancements that ease authentic language use. Taken together, these developments highlight the necessity of reevaluating traditional oral assessment frameworks such as Canale's, (1984) structured proficiency interview prioritizing approaches that both reflect communicative competence and utilize the precision and adaptability of AI-supported technologies.

2.5. Kinds of Tests

Tests vary according to their purposes and designs. Heaton, (1995), Hughes, (1996) and Brown, (2003) categorize tests into four basic types:

1- Achievement Test

This kind of tests is designed to measure what students have learned over a course or term, usually summative assessments administered at the end of instructional periods (Heaton, 1995).

2- Diagnostic Tests

The aim of such tests is to identify learners' strengths and weaknesses to inform targeted remedial instruction. Rivers (1968) describes diagnostic tests as instruments for demonstrating areas that require further practice before moving forward.

3- Proficiency Tests

This kind of tests assesses general language proficiency regardless of specific instruction or course completion. Examples include Cambridge's First Certificate and Proficiency Exams and Oxford EFL Tests (Hughes, 1996).

4- Language Aptitude Tests

It measures an individual's capacity to learn a foreign language and predict potential success. The Modern Language Aptitude Test (MLAT) is a widely used example (Brown, 2003; Carroll & Sapon, 1958).

2.6 Multiple Choice Items in Language Assessment

Hughes, (2003) argues that MCIs are efficient, reliable, and easy in scoring which led of being a dominant format in large-scale of testing. They are well known and effective tools in measuring different linguistic knowledge as (grammar, vocabulary, or listening comprehension) yet, they are criticized for their inability to assess productive performance (Fulcher, 2003). The validity of any test according to Bachman and Palmer (1996) depends on the alignment between the construct being measured and the tasks it utilizes. Since MCIs formats rely on recognition rather than actual and practical production, their construct validity as measurement of speaking competence is invalid. According to the Longman Dictionary of Language Teaching and Linguistics (Richards et al., 2018), the multiple choice item (MCI) format consists of a stem (a question or incomplete statement) followed by different alternatives, typically one correct answer and multiple distracters. Heaton (1995) explains that distracters are specifically designed to attract incorrect responses from less knowledgeable test-takers. Although MCIs are widely used as they are easy for scoring and administrating, their effectiveness is limited. Heaton (1995) cautions that MCIs often fail to assess speaking competence. Choosing one correct a set of options does not accurately reflect real-life language use, where responses are generated, not selected.

Seliger & Shohamy, (1989) and Brown, (1994) argue that MCIs formats can assess grammar and vocabulary knowledge but unable to measure learners' communicative skills such as pronunciation, stress, intonation, and interactive competence. Nambiar and Goon (1993) highlight the importance of paralinguistic features in oral communication, which MCIs cannot measure. Heaton, (1995) further criticizes MCI formats for de-contextualizing language and potentially hindering learning by suggesting that language can be used appropriately out of context. Moreover, Hughes (2003) enumerates several limitations:

1. They test recognition rather than production of (MCIs)
2. They allow guessing, which can skew scores.
3. They limit the scope of what can be tested.
4. They are difficult to construct effectively.
5. They may facilitate cheating.

According to Brown (2003), authentic assessment of speaking requires interactive tasks in which learners produce language spontaneously, a requirement that paper-based MCIs fail to meet. While MCI formats can provide indirect measurement of knowledge that triggers speaking skills (e.g., grammar, vocabulary), they cannot be considered a valid standardized tool for measuring learners' speaking competence. For Libyan secondary school teachers, an overreliance on MCIs risks reducing the construct of "speaking competence" to mere

recognition of linguistic forms, ignoring crucial aspects of oral communication. Future assessment practices should integrate performance-based tasks alongside objective items ensuring both validity and practicality.

2.7 Characteristics of a Good Test
Two fundamental traits define a good test: validity and reliability
4.1 Validity
Validity refers to the extent to which inferences drawn from test results are appropriate, meaningful, and aligned with the assessment purposes. Jaber (2018) emphasizes that a good test "must test what it is meant to test." Heaton (1995) similarly defines validity as "the extent to which a test measures what it is supposed to measure and nothing else." According to The Longman Dictionary of Language Teaching and Applied Linguistics (Richards, Platt, & Weber, 2018), validity is the degree to which a test fulfills its intended purpose. Contemporary research continues to highlight the critical importance of validity in language assessments where task authenticity impacts the meaningfulness of results (Li, Chen, & Wang, 2023).
4.2 Reliability
Reliability refers to consistency and dependability of a test. A reliable test produces similar results under consistent conditions regardless of the administrators or raters (Heaton, 1995). Richard, Platt, and Platt (2018) define reliability as the degree to which test results are reproducible over time or across different examiners. Recent advances in computer-assisted language testing have enhanced reliability especially in oral proficiency assessment through the use of automated scoring systems (Xu, Liu, & Chen, 2025).

The reviewed literature highlighted numerous studies regarding the use multiple-choice items format in assessing speaking proficiency and the comprehensive goals of the communicative pedagogy (Seliger & Shohamy, 1989; Brown, 1994; Heaton, 1995; Hughes, 2003). However, there remains an observable gap in research regarding the practical evaluation of speaking competence in secondary school contexts where MCIs are widely used. Previous research has constantly highlighted that MCIs can only measure grammar and vocabulary knowledge as they fail to capture essential oral communication skills such as fluency, pronunciation, intonation, stress, interactive competence, and paralinguistic features (Nambiar & Goon, 1993; Brown, 2003). Moreover, while studies have proposed authentic and task-based assessment methods (Brown, 2003; Canale, 1984), there is limited empirical studies in the Libyan secondary school context, where speaking instruction largely depends on MCIs. In addition, more recent advances in assessment, including AI-based tools, offer valuable alternatives that align more closely with communicative competence by emphasizing fluency, interaction, and authentic language use (Cha et al., 2024; Fang et al., 2025; Ma et al., 2025) but few studies have explored the alignment between these innovations and local classroom practices, teacher perceptions, and students' actual oral performance. In the Libyan secondary school context, however, an overreliance on MCIs risks reducing speaking to passive recognition of linguistic forms, adopting more integrative approaches becomes essential (Jaber, 2018). Taken together, these findings highlight the need for assessment practices that balance reliability, validity, and authenticity, ensuring that tests not only measure learners' knowledge but also reflect the real-world communicative abilities required for effective language use (Li, Chen, & Wang, 2023; Xu, Liu, & Chen, 2025). Thus, a significant gap remains in understanding the efficiency of MCIs to measure speaking proficiency in real classroom settings, the limitations they impose on teaching and learning, and the potential for integrating performance-based assessments to provide a valid and practical evaluation of oral competence in Libyan secondary schools.

This study addresses a critical gap in previous research by exploring how multiple-choice items (MCIs) measures speaking proficiency in Libyan secondary classrooms context. It also investigates teachers' perceptions and practices, in addition to evaluate the potential of task-based assessments to provide a more authentic and practical measurement of students' communicative competence. By linking assessment theory with classroom realities, this study sheds light on creating speaking assessments that maintain validity, reliability, and practical relevance.

3. Methodology
3.1 Method
Data collection was carried out through triangulation method, combining three instruments: document review, an oral test, and a teacher questionnaire.
3.2 Instrument
Three instruments were used for data collection:
    1- Document Review
Three sample speaking achievement tests of the first term were collected from three secondary schools based on MCI formats. They were analyzed using adopted criteria from Brown's (2003) taxonomy of micro- and macro-skills, which represent essential objectives in speaking assessment.

2- Oral Test

An oral proficiency test was designed to assess students' speaking abilities. Face-to-face interview items were adapted from Canale's (1984). Students engaged in the face-to-face interview and their performances were audio-recorded for subsequent analysis.

3- Questionnaire

The questionnaire was adapted from Heaton (1995) for exploring the limitations of MCIs in assessing speaking, Brown (2003) for micro- and macro-skills of speaking and principles of communicative assessment and Seliger & Shohamy (1989) for teachers' perspectives on language testing and assessment validity. It was designed to elicit secondary school English language teachers' perceptions of speaking assessment, with particular focus on the use of multiple-choice item (MCI) formats. Modifications were made to fit the Libyan secondary school context.

3.3 Participants

The study involved two groups of participants: students and teachers. Ten Libyan third year secondary school students participated in the oral test. All the students have been studying English for six years (three years at the preparatory level and three years at the secondary level). They all shared a relatively similar exposure to English language instruction. In addition to ten English language teachers who participated by completing a questionnaire. The participating teacher had 2-5 years experience in teaching English and speaking skills at the secondary school level.

3.4 Setting

The study was conducted in three secondary schools located in Sabha, Libya during the academic year 2024-2025. The schools were: Othman Ibn Affan, Al-Khansa'a, and Libya Libya Al_Hurra Secondary Schools.

4. Data Analysis
4.1 Document Review

Three samples of secondary school achievement tests of the first term were analyzed. Each test consisted of 10 multiple-choice items (MCIs). A breakdown of the content is provided below:

Sample 1: 5 items assessed vocabulary knowledge, 2 items tested grammar, 1 item addressed reading comprehension, and 2 items were based on listening comprehension.

Sample 2: 5 items tested vocabulary knowledge, 2 items addressed grammar, 2 item was based on listening comprehension, and 1 items focused on reading comprehension.

Sample 3: 4 items targeted vocabulary knowledge, 2 items tested grammar, and 1 item addressed reading comprehension, and 2 items were based on listening comprehension.

A cross-sample analysis of the tests' items revealed that vocabulary language knowledge included speaking focused items as (greetings, small talk, asking about hobbies, routines). None of the speaking items measured the micro- or macro-skills of speaking as classified by Brown (2003). This is due to the fact that such skills require oral production to be elicited and assessed, which cannot be achieved through written multiple-choice formats.

4.2 Questionnaire

Ten teachers participated in the study 8 females and 2 males. Responses from the ten English language teachers who were teaching English at Libyan secondary schools were summarized as follow:

Table 1: Results of the questionnaire

No	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Q1	Speaking should be assessed through interactive and communicative tasks.	0%	0%	0%	0%	100%
Q2	Multiple-choice items (MCIs) are effective for assessing speaking skills.	80%	20%	0%	0%	0%
Q3	MCIs fail to capture fluency, accuracy, and interaction in speaking.	0%	0%	0%	20%	80%
Q4	Oral tests (e.g., interviews, dialogues) provide more valid results than MCIs.	0%	0%	0%	60%	40%

Q5	MCIs are useful only as supplementary tools for speaking assessment.	0%	0%	0%	20%	80%
Q6	Students generally perform better in speaking when assessed through oral tasks.	0%	0%	10%	40%	50%
Q7	MCIs negatively affect students' motivation to speak English	0%	0%	0%	40%	60%
Q8	Performance-based speaking assessment is more aligned with real-life communication	0%	0%	0%	20%	80%
Q9	Teachers feel confident using oral tasks for assessing speaking	0%	0%	0%	50%	50%
Q10	Students prefer performance based speaking assessments over MCIs	0%	0%	10%	30%	60%

Results of the questionnaire

The results of the questionnaire indicate a strong consensus among teachers regarding the limitations of multiple-choice items (MCIs) for assessing speaking and the advantages of performance-based assessment.

Q1: All teachers (100%) strongly agreed that speaking should be assessed through interactive and communicative tasks, highlighting the importance of authentic language use in assessment.

Q2: A large majority of participants strongly disagreed (80%) and the others (20%) disagreed which indicated that MCIs are ineffective for assessing speaking skills, suggesting a clear perception that MCIs fail to capture critical aspects of oral proficiency.

Q3: Most teachers (80% strongly agree, 20% agree) confirmed that MCIs fail to assess fluency, accuracy, and interaction, reinforcing the need for tasks that require active speech production.

Q4: Teachers generally agreed (60% agree, 40% strongly agree) that oral tests such as interviews and dialogues provide more valid assessments than MCIs, demonstrating preference for performance-based evaluation.

Q5: 80% of teachers strongly agreed and 20% agreed that MCIs should only be used as supplementary tools, indicating that written formats are insufficient as primary assessment methods.

Q6: 50% strongly agreed and 40% agreed that students perform better in speaking when assessed through oral tasks supporting the effectiveness of interactive assessment formats. Meanwhile, 10% of teachers remained neutral, which suggested that although the majority recognize the benefits of oral assessments, some teachers are either unsure or see other factors influencing student performance.

Q7: Teachers reported that MCIs negatively affect student motivation (60% strongly agree, 40% agree), suggesting that written recognition tasks may hinder learner engagement.

Q8: 80% strongly agreed and 20% agreed that performance-based assessment better reflects real-life communication, emphasizing the relevance of authentic tasks.

Q9: Teachers felt confident in using oral tasks, with 50% agreeing and 50% strongly agreeing, indicating readiness to implement communicative assessment methods.

Q10: 10% strongly agreed and 30% agreed that students prefer performance-based assessments over MCIs, while 60% were neutral, showing alignment between teacher and student preferences for oral, interactive evaluation.

Overall, the findings clearly demonstrate that teachers in Libyan secondary schools favor performance-based and interactive speaking assessments over MCIs, citing issues of validity, engagement, and authenticity as key factors. This aligns with prior research highlighting the limitations of MCIs for measuring speaking competence (Heaton, 1995; Brown, 2003; Hughes, 2003).

4.3 Oral test results

An oral testing scale from (1-6) was adopted from (Heaton, 1995) to measure the scoring of the oral test. Interview questions were designed to assess speaking proficiency, aligned with the scoring rubric (1–6). These questions progress from simple to more complex in order to elicit differences in fluency, grammar, vocabulary, and communicative ability:

Interview questions for oral assessment

Section A: Personal and Everyday Topics (Level 1–2)

1. Can you introduce yourself and tell me a little about your family?
2. What do you usually do on weekends or during your free time?
3. Describe your daily routine.

Section B: Familiar Situations (Level 3–4)

4. Tell me about your favorite school subjects, and which one do you prefer?

5. Describe a recent event or activity you participated in at school or home.

6. What do you usually eat for breakfast, lunch, and dinner? Describe your favorite dish.

Section C: Opinion and Reasoning (Level 5–6)

7. What are the advantages and disadvantages of learning speaking in the school?

8. If you could travel anywhere in the world, where would you go and why?

9. How do you think technology affects the way people communicate today?

Section D: Problem-Solving / Hypothetical Situations (Level 5–6)

10. Imagine your friend is having a difficulty learning English. How would you help him?

11. What would you do if you lost your way in a new city? Describe your actions.

12. Discuss a school or community problem you would like to solve and explain your solution.

Table 2: Notes for scoring.

1-2	**Extreme difficulty, very limited vocabulary, frequent errors, minimal communication.**
3-4	Satisfactory/average verbal communication, some errors, needs repetition/rephrasing.
5-6	Fluent, accurate, good use of grammar, vocabulary, cohesive devices; able to handle more complex topics.

Table 3: Results of the oral test.

S. N	Rating	Communication Abilities
20%	4	Satisfactory verbal communication, limited number of errors of grammar, lexis and pronunciation. Good use of cohesive devices. Able to communicate on every day topics.
40%	3	Verbal communication is average; a speaker may experience some difficulty communicating with them. Repetition, rephrasing, are necessary. Errors in grammar, lexis, and pronunciation are noted.
30%	2	Much difficulty for a speaker to communicate with them. Their range of vocabulary is severely limited. A considerable number of grammatical errors are noted. A suitable use of hesitation fillers.
10%	1	Extreme difficulty in communication. Absence of fluency, slow production and responds with one word or two. Failure to express themselves and to be understood.

As shown in the table above, 20% of the students were rated 4 out of 6. These students demonstrated satisfactory verbal communication, with a limited number of errors in grammar and appropriate lexical choices for the given context. Minor pronunciation errors were observed, but they generally used cohesive devices effectively, allowing communication to be achieved with ease.

40% of the students received a rating of 3 out of 6. While they were able to communicate verbally, native speakers might experience some difficulty understanding them. Their speech often included repetitions, rephrasing, and restructuring, and lacked stylistic or expressive language features.

30% of the students were rated 2 out of 6. Communication with these students occurred with considerable difficulty due to their limited vocabulary and grammatical range. Their speech contained numerous phonological, lexical, and grammatical errors.

Finally, 10% of the students were rated 1 out of 6, reflecting extreme difficulty in communication. These students exhibited hesitancy, minimal responses (one or two words), frequent errors in pronunciation, grammar, and word selection, and were largely unable to express themselves or make their meaning understood.

5. Findings

The analysis of the speaking achievement tests revealed that none of the multiple-choice items directly assessed students' speaking abilities. Instead, these items focused on evaluating other competences, including reading comprehension, vocabulary, and grammar. Consequently, multiple-choice formats fail to provide an accurate measurement of oral communication skills. In contrast, the oral test clearly reflected students' speaking abilities, capturing essential skills such as fluency, organization and flow of ideas, and proper intonation patterns elements that can only be observed through actual oral production.

It was evident that multiple-choice items are inadequate for assessing speaking proficiency, as they merely measure students' knowledge of language components rather than their ability to communicate orally. The study also identified a significant lack of appropriate speaking materials in secondary schools. Although, the teachers claimed that they can use computer labs for English classes, they were under-equipped which limit opportunities for effective oral practice. They also experience technical difficulties, which restrict opportunities

for meaningful oral practice. In addition, the time and effort dedicated to teaching speaking were found to be insufficient, negatively impacting students' performance in oral tasks.

Data from the teacher questionnaire further reinforced these findings. Around 90% of the participating teachers expressed strong support for using activities that stimulate student speech, such as face to face interviews and dialogues, while rejecting the reliance on multiple-choice items for assessing speaking. The results also indicated that speaking instruction in Libyan secondary schools is largely exam driven, focusing on passing tests and obtaining grades rather than developing genuine communicative competence and the ability to speak appropriately in real-life contexts.

6. Conclusion

Although multiple-choice items (MCIs) are widely used in Libyan secondary educational context, they are invalid to capture the dynamic and interactive nature of speaking competence. MCIs are recognition-based tasks that mostly assess grammar and vocabulary, neglecting fluency, pronunciation, interaction, and pragmatic use of language. This limited construct validity, as noted by Bachman and Palmer (1996), risks creating a false impression of communicative ability. The results of this study is in consistent with research on dynamic and performance-based assessments (Siwathaworn & Wudthayagorn, 2018; Maryam & Jalil, 2020; Phung, Qin, & Nguyen, 2020; Al-Jamal & Al-Jamal, 2023; Ghazali & Ahmad, 2024). More authentic assessment approaches such as oral interviews, role plays, integrated performance tasks, and portfolio assessments can provide a direct measurement of speaking proficiency and better align with communicative competence principles. MCIs may serve as supplementary tools for testing underlying knowledge, but they should not substitute performance based assessments. For Libyan secondary schools, a mixed model approach that combines MCIs for underlying knowledge with oral tasks offers a practical yet valid solution to ensure reliable and authentic evaluation of students' speaking skills.

7. Limitations and suggestions for future research

Different limitations should be considered regarding the current study. First, this study focused on third-grade secondary students in Libya, which limited the applicability of the results. Different educational contexts or grade levels may yield different outcomes. Furthermore, the study did not incorporate different performance-based speaking tasks such as (role plays, or oral presentations) for comparison, which limited the opportunities to further validate the finding.

Several suggestions are offered, in light with limitations, for future research. First, future studies could include more participants from different regions, grade levels, or educational systems which could provide greater applicability and richer insights. Second, employing different mixed approaches that merge test results with qualitative data, such as classroom observations, or students perceptions could help expose both practical challenges and opportunities of speaking assessment. Third, technological innovations in language testing, such as computer-assisted or AI-based speaking assessments, deserve further investigation regarding the Libyan context. These tools may provide more reliable and authentic assessment of oral proficiency than traditional methods. Finally, longitudinal studies are also recommended to follow students' speaking development over a period of time, allowing researchers to better evaluate the strengths and limitations of different assessment approaches in capturing oral language development.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Al-Badri, M. (2018). Assessment beliefs and practices of Libyan secondary school English language teachers. *International Journal of Applied Linguistics & English Literature, 7*(6), 150–160.

Al-Jamal, D., & Al-Jamal, G. (2023). EFL teachers' speaking assessment practices in Jordanian secondary schools. *Frontiers in Education, 8*, Article 1192754.

Al-Zadma Al-Fourganee, R. (2018). *How secondary school Libyan teachers assess students' speaking skills* [Doctoral dissertation, University of Sunderland, UK].

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

Brown, H. D. (1994). *Principles of language learning and teaching* (3rd ed.). Prentice Hall Regents.

Brown, H. D. (1995). *Teaching by principles: An interactive approach to language pedagogy*. Prentice Hall Regents.

Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Pearson Education.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.

Burns, A., & Joyce, H. (1997). *Focus on speaking*. National Centre for English Language Teaching and Research.

Canale, M. (1984). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera (Ed.), *Communicative competence approaches to language proficiency assessment: Research and application* (pp. 107–122). Multilingual Matters.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158.

Chaney, A. L. (1988). *Teaching oral communication in grades K–8*. Allyn & Bacon.

Cohen, A. D. (1996). Developing pragmatic ability: Insights from the study of speech acts. *Studies in Second Language Acquisition, 18*(2), 253–267.

Dostal, H. (2007). Speaking skills: Planning, fluency, and coherence. *Language Learning Journal, 35*(1), 43–50.

Elabbar, A. (2011). *An investigation of influences affecting Libyan English as Foreign Language university teachers in adapting communicative language teaching in their classrooms* [Doctoral dissertation, University of Glasgow].

Fulcher, G. (2003). *Testing second language speaking*. Pearson Education.

Gebhard, J. G. (1996). *Teaching English as a foreign or second language: A self-development and methodology guide*. University of Michigan Press.

Ghazali, M., & Ahmad, R. (2024). Student-led assessment in EFL speaking classrooms. *Language Teaching Research*. Advance online publication.

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford University Press.

Hassan, A. (2019). Analyzing classroom practices in Libyan preparatory schools: The role of multiple-choice items in assessment. *Faculty of Arts Journal, University of Tripoli, 5*(2), 45–60.

Heaton, J. B. (1995). *Writing English language tests* (2nd ed.). Longman.

Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.

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

Jaber, F. (2018). Principles of language testing. *International Journal of English Language and Linguistics Research, 6*(5), 1–8.

Knight, B. (1992). Assessing speaking skills: A workshop for teacher development. *ELT Journal, 46*(3), 294–302.

Li, X., Chen, L., & Wang, Y. (2023). Task authenticity and validity in speaking assessment: A review of recent research. *Language Testing, 40*(1), 97–120.

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

Maryam, M., & Jalil, A. (2020). Dynamic assessment and its effects on Iranian EFL learners' speaking skills. *International Journal of Language Studies, 14*(2), 67–84.

McNamara, T. (2000). *Language testing*. Oxford University Press.

Morozova, Y. (2013). Methods of enhancing speaking skills of EFL learners. *English Language Teaching, 6*(3), 18–27.

Nambiar, R. M. K., & Goon, P. (1993). The role of paralinguistic features in communication. *RELC Journal, 24*(1), 1–18.

Nunan, D. (1999). *Second language teaching and learning*. Heinle & Heinle.

Nunan, D. (2003). *Practical English language teaching*. McGraw Hill.

O'Sullivan, B. (2012). Assessing speaking. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 234–241). Cambridge University Press.

Phung, L., Qin, X., & Nguyen, T. (2020). A narrative review of speaking assessment research: Issues, challenges, and future directions. *Frontiers in Psychology, 11*, Article 330.

Richards, J. C., Platt, J., & Weber, H. (2018). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods*. Oxford University Press.

Solcova, P. (2011). Teaching speaking skills. *Brno Studies in English, 37*(2), 145–158.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Thornbury, S. (2005). *How to teach speaking*. Pearson Education.

Wang, X., Zhang, Y., & Li, X. (2023). Evaluating an AI speaking assessment tool: Score accuracy and learner perceptions. *Computer Assisted Language Learning, 36*(5), 1234–1256.

Xu, Y., Liu, Z., & Chen, H. (2025). AI-enhanced assessment of EFL speaking skills using deep learning. *Journal of Language Technology Research, 7*(1), 45–62.

Zhang, W., & Lu, S. (2021). The effect of strategic planning on EFL learners' oral fluency. *TESOL Quarterly, 55*(3), 560–578.

Zhang, Y., & Lu, X. (2021). Effects of strategic planning on L2 speech production: A meta-analysis. *Language Teaching Research, 25*(4), 558–580.

Teachers' Questionnaire
Section A: Background Information
1. Gender: ☐ Male ☐ Female
2. Years of teaching English: ☐ 1–2 ☐ 3–5 ☐ 6–10 ☐ 10+
Section B: Teachers' Perceptions of Speaking Assessment
(5-point Likert scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Table 1: Results of the questionnaire.

No	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Q1	Speaking should be assessed through interactive and communicative tasks.					
Q2	Multiple-choice items (MCIs) are effective for assessing speaking skills.					
Q3	MCIs fail to capture fluency, accuracy, and interaction in speaking.					
Q4	Oral tests (e.g., interviews, dialogues) provide more valid results than MCIs.					
Q5	MCIs are useful only as supplementary tools for speaking assessment.					
Q6	Students generally perform better in speaking when assessed through oral tasks.					
Q7	MCIs negatively affect students' motivation to speak English					
Q8	Performance-based speaking assessment is more aligned with real-life communication					
Q9	Teachers feel confident using oral tasks for assessing speaking					
Q10	Students prefer performance based speaking assessments over MCIs					

Appendix II

Interview Questions for Oral Assessment
Section A: Personal and Everyday Topics (Level 1–2)
1. Can you introduce yourself and tell me a little about your family?
2. What do you usually do on weekends or during your free time?
3. Describe your daily routine.
Section B: Familiar Situations (Level 3–4)
4. Tell me about your favorite school subjects and which one do you prefer?
5. Describe a recent event or activity you participated in at school or home.
6. What do you usually eat for breakfast, lunch, and dinner? Describe your favorite dish.
Section C: Opinion and Reasoning (Level 5–6)
7. What are the advantages and disadvantages of learning speaking in the school?
8. If you could travel anywhere in the world, where would you go and why?
9. How do you think technology affects the way people communicate today?
Section D: Problem-Solving / Hypothetical Situations (Level 5–6)
10. Imagine your friend is having a difficulty learning English. How would you help him?
11. What would you do if you lost your way in a new city? Describe your actions.
12. Discuss a school or community problem you would like to solve and explain your solution.

Appendix III
Secondary Three Achievement Tests (10 MCIs each) –Libya
Sample 1
1. "Conversation" means:
   a) Reading a book
   b) Talking with someone
   c) Writing a letter
   d) Listening to music
2. Choose the correct sentence: "I ……… to school every day."
   a) goes
   b) go
   c) going
   d) gone
3. Which phrase is suitable for greeting a teacher in the morning?
   a) Good night
   b) Good morning
   c) See you
   d) How old are you
4. Complete the dialogue:
   A: "How was your weekend?"
   B: "………"
   a) It was fun
   b) I like pizza
   c) Stop it
   d) What is your name?
5. Which question would you ask to find out someone's hobby?
   a) What time do you wake up?
   b) Where do you live?
   c) What do you like to do in your free time?
   d) How old are you?
6. Reading passage: "Amal studies English every day after school."
   Question: How could you ask Amal about her English study routine?
   a) What do you do after school?
   b) Where is the bus?
   c) How old are you?
   d) What is your favorite subject?
7. Past tense: "Yesterday, we ……… football in the park."
   a) play
   b) played
   c) playing
   d) plays
8. Listening: "Ali usually wakes up at 7:00 a.m."
   Question: What question could you ask Ali to find out his wake-up time?
   a) What time do you wake up?
   b) Where do you live?
   c) How old are you?
   d) Do you like football?
9. Which sentence shows polite clarification?
   a) What do you mean?
   b) Speak louder!
   c) Stop!
   d) Never mind
10. Listening: "Sara likes to read books every evening."
    Question: How could you start a conversation about Sara's hobby?
    a) What books do you like to read?
    b) How old are you?
    c) Where do you live?
    d) Stop reading
Sample 2
1. "Dialogue" means:

a) Reading a story
b) Talking with someone
c) Writing notes
d) Listening to music
2. Choose the correct sentence: "She ……… to school by bus every day."
  a) go
  b) goes
  c) going
  d) gone
3. Which phrase would you use to politely greet a classmate in the afternoon?
  a) Good afternoon
  b) Good night
  c) See you
  d) How old are you
4. Complete the dialogue:
  A: "Did you enjoy the school trip?"
  B: "………"
  a) Yes, it was amazing
  b) I like pizza
  c) Stop it
  d) What is your name?
5. Which question asks about someone's favorite subject?
  a) What is your favorite subject?
  b) Where do you live?
  c) How old are you?
  d) What time do you wake up?
6. Reading passage: "Omar studies English and Maths every day after school."
  Question: How could you ask Omar about his study routine?
  a) What do you do after school?
  b) How old are you?
  c) Where is the library?
  d) What do you like to eat?
7. Past tense: "Last weekend, we ……… swimming at the beach."
  a) go
  b) went
  c) going
  d) goes
8. Listening: "Huda usually wakes up at 6:30 a.m."
  Question: What question could you ask Huda to find out her wake-up time?
  a) What time do you wake up?
  b) Where do you live?
  c) How old are you?
  d) Do you like reading?
9. Which sentence is a polite way to ask someone to repeat something?
  a) Could you repeat that, please?
  b) Speak louder!
  c) Stop!
  d) Never mind
10. Listening: "Khaled likes to play football every evening."
  Question: How could you start a conversation about Khaled's hobby?
  a) What do you like to do in the evening?
  b) How old are you?
  c) Where do you live?
  d) Stop playing

Sample 3
1. "Small talk" means:
  a) Talking briefly about everyday topics
  b) Writing a story
  c) Reading a book
  d) Listening to music

2. Choose the correct sentence: "They ……… English every day."
   a) speaks
   b) speak
   c) speaking
   d) spoke

3. Which phrase is suitable for starting a conversation with a new friend?
   a) Nice to meet you
   b) Goodbye
   c) Stop
   d) How old are you

4. Complete the dialogue:
   A: "What did you do yesterday?"
   B: "………"
   a) I went to the park
   b) I like pizza
   c) Stop it
   d) How old are you?

5. Which question asks about someone's favorite food?
   a) What is your favorite food?
   b) Where do you live?
   c) How old are you?
   d) What time do you wake up?

6. Reading passage: "Lina studies English, Arabic, and Maths after school."
   Question: How could you ask Lina about her daily study routine?
   a) What do you study after school?
   b) How old are you?
   c) Where is the library?
   d) What do you like to eat?

7. Past tense: "Yesterday, we ……… a movie at home."
   a) watch
   b) watched
   c) watching
   d) watches

8. Listening: "Mohammed usually wakes up at 6:00 a.m."
   Question: What question could you ask Mohammed to find out his wake-up time?
   a) What time do you wake up?
   b) Where do you live?
   c) How old are you?
   d) Do you like reading?

9. Which sentence politely asks for repetition?
   a) Could you say that again, please?
   b) Speak louder!
   c) Stop!
   d) Never mind

10. Listening: "Salma likes to draw in her free time."
    Question: How could you start a conversation about Salma's hobby?
    a) What do you like to draw?
    b) How old are you?
    c) Where do you live?
    d) Stop drawing