

Received on (23-11-2024) Accepted on (22-12-2024)

<https://doi.org/10.33976/IUGJEPS.33.1/2025/6>

## Training Program in Authentic Assessment for Enhancing English Language Teachers' Assessment Literacy and its impact on improving their Students' Speaking Skills.

Samah I. Eljaish<sup>\*1</sup>, Dr. Akram S.I Habeeb<sup>\*2</sup>, Prof. Ibrahim H. Alastal<sup>\*3</sup>

Ministry of Education- Palestine Islamic University<sup>\*1,2,3</sup>

\*Corresponding Author: [sjaish@hotmail.com](mailto:sjaish@hotmail.com) , [ahabeeb@iugaza.edu.ps](mailto:ahabeeb@iugaza.edu.ps) , [ihastal@iugaza.edu.ps](mailto:ihastal@iugaza.edu.ps)

### Abstract:

#### Introduction:

Authentic assessment can serve as a powerful tool for assessing students' competencies in the context of global educational reforms.

**Aim:** This study aimed to assess the needs of English language teachers in order to design a training program focused on authentic assessment, and to evaluate the effectiveness of the program in improving teacher performance and enhancing students' speaking skills.

**Methodology:** This study employed a mixed-methods approach, combining quantitative and qualitative research. A one-group experimental design (pre-post-test) was used. The needs assessment sample consisted of 365 teachers, while the sample for the training program included 20 English language teachers and 30 students. Data were collected using a self-administered questionnaire and observation checklist, covering socio-demographic information, strategies, and tools for authentic assessment. SPSS version 25 was used for statistical analysis.

**Results:** The results showed a statistically significant improvement in the use of authentic assessment strategies (mean score of 2.64 in the pretest, increasing to 4.56 in the posttest, with a 42.1% change,  $p < 0.001$ ). The greatest improvement was observed in the use of observation strategy (mean score of 2.27 in the pretest, increasing to 4.66 in the posttest, with a 51.2% change). Additionally, there was a significant improvement in the use of all authentic assessment tools (mean score of 2.63 in the pretest, increasing to 4.39 in the posttest, with a 40% change). Female teachers and those with 11-15 years of experience showed significantly better use of authentic assessment strategies after the training ( $p < 0.001$ ). Observations of students revealed a statistically significant improvement in speaking skills, with a percentage change of 32.1% ( $p < 0.001$ ).

**Conclusion:** The training program was effective in enhancing teachers' use of authentic assessment strategies and in improving students' speaking skills.

**Recommendations:** The study recommended the implementation of the training program for all English language teachers, and that the authentic assessment needs to be frequently applied.

**Key words:** Authentic Assessment Strategies, English Language Teachers, Training Program, Governmental Schools, Gaza Strip, Palestine.

### برنامج تدريبي في التقييم الأصيل لتنمية الممارسات التقييمية لمعلمي اللغة الإنجليزية وأثره في تطوير مهارة التحدث لدى طلابهم

سماح إسماعيل الجيش<sup>1</sup>، د. أكرم صبحي حبيب<sup>2</sup>، أ.د. إبراهيم حامد الأسطل<sup>3</sup>

المخلص:

هدف الدراسة: هدفت الدراسة إلى تحديد احتياجات معلمي اللغة الإنجليزية لإعداد برنامج تدريبي قائم على التقييم الواقعي، وتقييم فعالية هذا البرنامج في تحسين أداء المعلمين، ومدى تأثيره على تطوير مهارة التحدث لدى الطلبة.

الطريقة والإجراءات: استخدمت الدراسة المنهج المختلط الكمي والنوعي. تم تطبيق المنهج التجريبي (مجموعة واحدة) باستخدام الاختبار القبلي والاختبار البعدي. لمعرفة احتياجات المعلمين، تم توزيع استبانات على 365 معلماً ومعلمة من معلمي اللغة الإنجليزية في جميع محافظات غزة. بناءً على نتائج الاستبانات، تم إعداد برنامج تدريبي في التقييم الواقعي، ثم تم تطبيق البرنامج على 20 معلماً ومعلمة. ولتقييم تأثير تطبيق ما تعلمه المعلمون على أداء الطلبة في مهارة التحدث، تم إجراء ملاحظات على 30 طالباً وطالبة. تم جمع البيانات باستخدام استبانة لاستراتيجيات وأدوات التقييم الواقعي، وقائمة ملاحظة صفية للمعلمين، وقائمة ملاحظة صفية للطلبة. وتحليل البيانات، تم استخدام برنامج الرزم الإحصائية (النسخة 25).

النتائج: أظهرت النتائج وجود تحسن ملحوظ ودال إحصائياً في استخدام استراتيجيات التقييم الواقعي، حيث ارتفع متوسط الدرجة من 2.64 في الاختبار القبلي إلى 4.56 في الاختبار البعدي بنسبة تغيير بلغت 42.1%. وكانت أعلى نسبة تحسن في استراتيجية الملاحظة، حيث ارتفع المتوسط من 2.27 إلى 4.66 بنسبة تغيير 51.2%. كما أظهرت النتائج تحسناً ملحوظاً ودالاً إحصائياً في استخدام أدوات التقييم الواقعي، حيث ارتفع المتوسط من 2.63 إلى 4.39 بنسبة تغيير بلغت 40%. كما أظهرت النتائج أن المعلمات الإناث والمعلمين ذوي الخبرة بين 11 - 15 سنة قد استخدموا استراتيجيات التقييم الواقعي بشكل أفضل بعد البرنامج التدريبي مقارنة بأقرانهم ( $\alpha < 0.001$ ). كما تبين وجود تحسن ملحوظ ودال إحصائياً في مهارات التحدث باللغة الإنجليزية لدى الطلبة بعد البرنامج التدريبي، حيث بلغت نسبة التغيير 32.1%.

الاستنتاجات: أثبتت الدراسة أن البرنامج التدريبي كان فعالاً في تحسين استخدام معلمي اللغة الإنجليزية لاستراتيجيات التقييم الواقعي، وأسهم في تحسين مهارات التحدث باللغة الإنجليزية لدى الطلبة.

التوصيات: أوصت الدراسة بضرورة تطبيق البرنامج التدريبي على جميع معلمي اللغة الإنجليزية، واستخدام التقييم الواقعي بشكل مستمر في تقييم أداء الطلاب.

الكلمات المفتاحية: استراتيجيات التقييم الواقعي، معلمي اللغة الإنجليزية، البرنامج التدريبي، المدارس الحكومية، قطاع غزة، فلسطين.

## **Introduction:**

Assessment is an essential component of the English as a Foreign Language (EFL) curriculum. In contemporary language education, there has been a growing shift from traditional testing methods toward adopting authentic assessment practices (Ortega & Minchala, 2017). Evaluating students' performance in the language classroom is one of the most critical responsibilities of English language teachers (ELTs), as assessment plays a significant role in enhancing learning outcomes. To ensure effective assessment, it must be both appropriate and authentic. Teachers need to select assessment practices that align with specific learning contexts to best support student learning (Gallavan, 2009). The quality of assessment is closely linked to students' academic performance, emphasizing its importance in the educational process. Research suggests that teachers spend up to one-third of their professional time engaged in assessment-related activities. However, many teachers feel underprepared due to insufficient focus on assessment literacy during their training (Stiggins & Conklin, 2019).

Authentic assessment is increasingly recognized as a valuable tool for evaluating students' competencies within the framework of global educational reforms. This approach involves using diverse techniques to gather information about students' progress and learning achievements. Beyond determining whether learning objectives have been met, authentic assessment aims to provide ongoing and comprehensive insights into students' learning outcomes (Zaim et al., 2020).

Despite its potential benefits, studies reveal varied levels of awareness and understanding of authentic assessment among educators. Trisanti (2014) highlighted limited knowledge and understanding of how to implement authentic assessment effectively. Similarly, Maulidhawati et al. (2021) found that while authentic assessment positively impacts students' English learning achievement, teachers often face challenges in its implementation due to insufficient knowledge and training.

In this study, the researcher built and implemented a training program in authentic assessment to meet the teachers' needs.

## **Research problem:**

An intensive review of the educational literature showed that there is a shortage of the empirical research studies which concerns with ELT practices, especially assessment skills which teachers should be equipped with (Furtak et al., 2013). Research in the field of authentic assessment reflected low awareness and implementation of authentic assessment by teachers. In this regard, Yamtim & Wongwanich (2014) found that most of the teachers had low level of classroom assessment literacy, and suggested that improving the classroom assessment literacy should emphasize cooperative learning and teamwork, with knowledgeable persons acting as mentors or coaches who offer advice during teaching practices, while Hussain et al. (2019) stated that lack of training in assessment was a major challenge for teachers. In addition, Zaim et al. (2020) reported that the problem faced by teachers in implementing authentic assessment is the lack of teachers' experience, and it takes a lot of time and effort to administer student grades and subjectivity in giving the final results. From the researcher experience as ELT for more than 15 years and working in the assessment and evaluation department in the Ministry of Education (MoE), the researcher noticed that there is a pressing need for improving ELT skills in assessment. This study is a responsive act to the actual need to enhance teachers' classroom practices in assessment. In addition, it will help in raising their awareness of the importance of using their own assessment competencies in evaluating their students' achievements.

## **Research question (s):**

What is the effectiveness of the training program based on authentic assessment in enhancing English language teachers' performance?

## **Aim of the study:**

This study aimed to assess ELTs needs in order to build a training program in authentic assessment, and to evaluate the effectiveness of the training program in enhancing English language teacher performance and its impact on their students speaking skills.

### **Literature review:**

Authentic assessment is an approach to evaluation that focuses on assessing students' ability to apply their knowledge and skills in real-world contexts, rather than just measuring their ability to recall information (Nguyen & Phan, 2020). In English language teaching, learning should meet students' needs so that they would be able to apply their knowledge into their own lives. Thus, final outcome, or final evaluation, accordingly encompasses authenticity (Idham et al., 2015).

### **The concept of authentic assessment**

Authentic assessment involves evaluating learning through tasks that reflect real-world scenarios, requiring students to demonstrate their knowledge and skills in meaningful contexts. These assessments are designed to measure a learner's ability to apply their learning to practical, real-life situations (Manchester Metropolitan University, 2023). In essence, authentic assessment tasks challenge learners to showcase their performance in ways that meaningfully apply knowledge and skills to real-world contexts.

Similarly, Barnawi (2022) highlights that authentic assessment is closely linked to real-world evaluation. It emphasizes assessing students' abilities to demonstrate learning outcomes that extend beyond the classroom, reflecting skills applicable in everyday life. By incorporating authentic assessment, students not only apply their knowledge and skills in practice but also develop competencies relevant to job-related tasks. Authentic assessments provide opportunities for students to engage in tasks within real or simulated environments, demonstrating their mastery of relevant knowledge, skills, and competencies through direct application.

### **Benefits of authentic assessment**

Authentic assessment proves valuable in most learning contexts but is particularly significant in inclusive language learning settings and across all age groups. This approach offers an alternative to standardized testing, enabling instructors to adopt more holistic and reliable methods. An analysis of various authentic assessment techniques suggests that alternative testing methods are especially beneficial in inclusive second language learning classrooms (Reynisdóttir, 2016).

The benefits of authentic assessment for students are numerous. It fosters creative and critical thinking by exposing learners to unpredictable scenarios and complex environments. It also engages students in meaningful tasks, emphasizing the practical value of what they are learning. Moreover, authentic assessment allows students to integrate their lateral skills and competencies with subject-specific expertise, encouraging a well-rounded approach to learning. It further enhances self-efficacy and work-readiness by immersing students in real-world situations. For educators, authentic assessment offers direct evidence of students' competencies, practical skills, and ability to apply theoretical knowledge in practice (The University of Melbourne, 2022).

### **Use of authentic assessment for English language skills**

Assessment is a crucial and integral part of the teaching-learning process, helping to determine whether the goals of teaching and learning are being met. One unique method of assessing students' reading comprehension is through authentic assessment, which demands professional skills from teachers, as it requires independent evaluation and interpretation of student performance. This method takes time and careful planning to implement effectively. However, many teachers feel unprepared and unsupported when it comes to using authentic assessment. This study highlights some of the challenges teachers face in implementing authentic assessment. Based on observations and interviews, the researcher identified

several difficulties teachers encounter in preparing, applying, and implementing authentic assessments for students' reading (Ekawati, 2017).

Refnita (2013), identifies three forms of adaptation: sorting, adding, and modifying the material to create newer, student-friendly material. In addition, McDonough et al. (2013) pointed out five techniques for adjusting the model or material (1) adding; including expanding and extending, (2) deleting; including subtracting and abridging, (3) modifying; including rewriting and restructuring, (4) simplifying, and (5) reordering. In connection with the above information, it is important to adapt the authentic assessment for speaking skills. It is therefore necessary to carry out a needs analysis. By conducting a needs analysis, an authentic assessment that is appropriate and applicable to students can be developed (Zaim & Refnaldi, 2016).

### **Strategies to implement authentic assessment**

#### **Paper and pencil strategy**

Paper-and-pencil (PP) testing was once the dominant method of delivering assessments. Its advantages include students' familiarity with the format, the opportunity to foster social connections among learners, immunity to technical issues, and affordability. Paper exams are particularly cost-effective in situations where organizations lack access to professional assessment platforms or learning management systems (Stephens, 2021).

Despite advancements in assessment methods, research by Al-Zoubi (2019) indicates that paper-and-pencil testing remains the most commonly used approach for evaluating student learning. However, Azim and Khan (2012) reported positive shifts in perceptions and practices when traditional paper-and-pencil tests were replaced with authentic assessments. Their study highlighted increased engagement from both teachers and students in the teaching and learning process, alongside significant improvements in students' higher-order thinking skills. Through authentic assessments, students became active participants in planning, gathering information, and sharing their knowledge with the broader community.

#### **Performance-based assessment strategy**

Performance-based assessment (PBA) requires students to actively participate in tasks, enabling the evaluation of a broad spectrum of learning, reasoning, and communication skills. As a type of authentic assessment, PBA involves students performing tasks directly, providing educators with a more accurate and practical application of skills compared to traditional paper-and-pencil tests (Herrera et al., 2013). PBA is characterized by its high level of authenticity, as it closely mirrors real-world scenarios, ensuring that the assessment aligns with its intended goals (Alhassan et al., 2020). This approach allows students to demonstrate productive skills while giving educators valuable feedback on the effectiveness of their teaching methods, helping them refine their strategies and make necessary adjustments (Lund & Kirk, 2019).

PBA also fosters student autonomy by encouraging learners to engage with real-life tasks (Alhassan et al., 2020). The validity of an assessment, crucial for its effectiveness, depends on its ability to measure the intended skills accurately. Teachers can design valid assessments that align with learning objectives to evaluate communication skills effectively. For instance, creating activities that directly reflect course objectives ensures a reliable measure of student performance. Research by Gallardo (2020) revealed that PBA emphasizes student engagement with specific tasks to demonstrate their knowledge and skills. These tasks are typically student-centered and reflect real-world issues (Koné, 2021). Furthermore, a study by AlKhateeb (2018) found a statistically significant relationship between outcome test scores and global self-efficacy measures, with results favoring the experimental group, highlighting the positive impact of PBA on student achievement and confidence.

### **Observation strategy**

Classroom observation is a key strategy for studying and assessing teaching and learning processes. The data collected during classroom observations are typically supported by evidence from other sources, such as student performance, teaching materials, administrative reviews, and teacher self-assessments. This makes classroom observation one of the most direct methods for gathering insights into a teacher's instructional practices (MacDonald, 2016).

MacDonald (2016) emphasizes the importance of having a clear focus during observations. Without a guiding purpose, it becomes challenging to draw meaningful comparisons or conclusions. Defining the purpose of the observation ensures that the most relevant aspects of teaching and learning are observed deliberately and attentively. Importantly, the focus of classroom observation should extend beyond the teacher's performance alone. Instead, it should prioritize identifying patterns of teaching and learning through teacher-student interactions and student engagement in response to the teacher's actions.

In this context, observers are encouraged to adopt a broader perspective, akin to widening the angle of a camera lens, to capture the full range of behaviors and interactions in the classroom. This approach provides a more comprehensive understanding of how student learning is influenced by the teacher's decisions, communication, and actions, ultimately leading to more effective insights for improving teaching practices.

### **Communication strategy**

This strategy involves teachers gathering information by engaging directly with students in the classroom to assess their progress, as well as to evaluate improvements in their thinking and problem-solving styles. A variety of activities fall under this strategy, including interviews, which are planned meetings between teachers and students. These interviews provide teachers with valuable insights into students' ideas, attitudes toward specific topics, and overall progress. Additionally, they help evaluate students' work on projects, research, or similar tasks by discussing their current status and identifying the next steps needed to enhance learning outcomes. Another key activity within this communication strategy is the use of questions and answers. Teachers pose direct questions to students to monitor their progress, gain insights into their thought processes, and understand their problem-solving approaches (Al-Najjar, 2018).

In summary, effective communication is essential for monitoring and evaluation. It involves collecting, analyzing, and using data to inform decision-making and improve outcomes. However, without clear and consistent communication, the data collected may be misinterpreted, improperly shared, or not utilized effectively, diminishing its impact on decision-making and learning enhancement.

### **Self-assessment strategy**

While learning styles may vary, review remains a fundamental aspect of the language learning process. Revisiting material over time is essential for reinforcing knowledge, preparing for new concepts, and excelling in exams. Research shows that students often forget key concepts once they leave the classroom, making regular revision critical for deepening their understanding (Bjork & Kroll, 2015; Rosenshine, 2012). To support sustainable learning, students should actively engage in the feedback process. This includes not only interpreting and acting on teacher feedback but also contributing feedback to themselves and their peers. A study conducted with Chinese EFL (English-as-a-Foreign-Language) students investigated their perceptions of feedback practices, focusing on teacher feedback, peer review, and self-assessment. Data gathered from focus group interviews and classroom observations revealed that while students valued teacher feedback for its role in their progress, they placed less importance on feedback from peers or self-evaluation.

Three key reasons emerged for students' hesitance to embrace peer and self-feedback: a lack of understanding of their roles in the feedback process, doubts about their ability to provide meaningful feedback, and emotional or relational concerns when engaging with peers during feedback activities.

These findings underscore the importance of building students' feedback literacy. By helping students understand and utilize feedback from multiple sources, educators can enhance their learning and foster long-term academic growth (Wu et al., 2021).

### **Tools of authentic assessment**

#### **Rubrics**

In education field, rubrics are scoring systems that assist and act as a guide to judge different constructs, including the quality of work done by students, academic performance, and educational resources (Porcello & Hsi, 2013). Rubrics are assessment tools created to help identify and evaluate the qualitative differences in student performance (Jönsson & Panadero, 2017). A rubric outlines the criteria for scoring and defines the various levels of quality (Arter & Chappuis, 2006). In essence, rubrics are purposely designed to make clear the requirements for achieving different levels of quality.

#### **Checklists**

A checklist is an observational tool that the teacher needs to describe the learning situations and behaviors being assessed. When used regularly throughout the school year, growth checklists provide an interesting picture of students' progress toward the desired skills. Most checklists have a yes/no format to indicate specific student criteria and can be used to record individual, group, or whole class observations. The development of the checklist should also be considered. The checklist should be accurate. Cunningsworth (1995) states: "It is important to limit the number of criteria used, the number of questions to limit the number of criteria that are used, to a manageable number, if there is a risk that we are drowning in a sea of details". If the checklist is accurate and concise when used for audit purposes, it can be quick and save a lot of time and money.

#### **Rating scale**

A Rating Scale (RS) is an assessment tool used to evaluate the performance of tasks, skills, procedures, processes, qualities, quantities, or final products, such as reports or drawings. These are assessed at specific levels within a defined range. RSs serve as guidelines that clearly outline performance expectations and proficiency levels (Gezie et al., 2012) and are used to assess student work (Dawson, 2017). Unlike checklists, which simply evaluate whether a specific step, property, or action is present or absent, RSs assess more than just the presence or absence of these elements. RSs measure the degree of accomplishment in a task, indicating the frequency or quality of behaviors, skills, and strategies demonstrated by the student, rather than providing a simple yes or no answer. RSs outline criteria and offer three or four response options that describe the quality or frequency of student work, using descriptive bands to represent varying levels of competence (Kabir, 2012).

#### **Learning log**

A Learning Log (LL) is a personal record or journal, typically informal, where students document their learning experiences. It is a unique document for each student and does not have a right or wrong format. LLs help students reflect on, plan, and track their learning progress (Rizvi, 2014). They provide insights into the learning processes and strategies students use, allowing teachers to assess these processes and understand students' attitudes and motivations toward learning English (Brown, 1998).

By providing a deeper understanding of students' strategies, LLs help them solve problems and overcome challenges more effectively. These journals encourage students to recognize what they have learned and identify areas for improvement, making them more engaged and self-aware in their learning process (Teaching Today, 2011).

#### **Anecdotal record**

Among various observation tools, anecdotal record-keeping is considered one of the most effective for tracking students' progress. The American Association of School Administrators defines an anecdotal record as, "a written record kept in a positive tone of a child's progress based on milestones peculiar to

that child's physical, social, economic, aesthetic, and cognitive development" (AASA, 1992). Anecdotal records provide a longitudinal and qualitative account of a student's behavioral changes, offering insights into their development over time. Baker et al. (2000) note that observational notes allow teachers to capture a wide range of authentic experiences, including unintended outcomes of literacy development. These notes record both objective and subjective information, such as levels of engagement, curiosity, and motivation, offering a comprehensive view of a student's learning process. Anecdotal records document significant events, whether in the classroom or outside, that highlight a student's behavior, thinking, skills, and personal qualities.

### **Previous studies:**

Several studies have highlighted significant variability in teachers' understanding and implementation of authentic assessment. Many teachers struggle to grasp the concept fully or to apply it effectively in their teaching practices. For instance, a study conducted in Batusangkar, Indonesia, revealed that English teachers utilized authentic assessment for speaking but failed to adequately address students' needs or knowledge levels. Furthermore, the assessments lacked specific scoring rubrics to evaluate student performance comprehensively. This finding aligns with Zaim (2017), who observed that only a small number of teachers successfully implement authentic assessment.

A survey by Rusilowati (2013) of 20 teachers found that 87% faced difficulties in creating authentic assessments, 87% struggled to understand assessment methods, and 70% encountered challenges in developing observation instruments. Similarly, an observational study in primary schools in Medan, Indonesia, revealed that nearly all teachers had difficulty assessing students, particularly in evaluating character development. Across five primary schools, teachers reported challenges in using authentic assessments to measure students' achievement of specific character values (Setiawan & Hadikusuma, 2015).

Aziz et al. (2020) identified several key challenges teachers face when preparing for authentic assessments, grouped into five themes: lack of support from school administration and parents, excessive documentation requirements, heavy teaching workloads, frequent weekend school activities, and insufficient training in authentic assessment practices. In another study conducted in Indonesia, teachers implementing authentic assessment for speaking reported encountering obstacles such as limited time to assess all students effectively, leading to suboptimal evaluations. Teachers also struggled to categorize speaking rubrics according to different proficiency levels (Elsara & Maiwen, 2019). In Ghana, a descriptive study investigated the impact of authentic assessment strategies on secondary school students' performance. Teachers employed various strategies, including portfolios, project work, rubrics, observation, self and peer assessments, storytelling, and demonstrations. Despite this variety, numerous challenges hindered the effective use of authentic assessment. These included difficulties in building student confidence for oral presentations, managing large class sizes, completing performance tasks, adhering to time constraints during meetings, and addressing the increased demands for time and resources (Butakor & Ceasar, 2021).

## **Materials and Methods**

### **Study design**

The study utilized mixed method (quantitative and qualitative). The researcher used a one-group experimental design (pre-post-test). A pre-post-test study measures the occurrence of an outcome before and after a particular intervention is implemented.

### **Population of the study**

The target group of the study consisted of all schools (fifth to tenth grade) EL teachers at governmental schools in Gaza Strip, and their total number is 769.

### **Sample and sampling method**

The sample of the study for the need assessment was 365 teachers selected by stratified, proportional, random sampling method. The sample for the training program consisted of 20 EL teachers and 30 of their students who were assessed before and after participating in the training program which took approximately two months (30 training hours).

### **Setting of the study**

The study was conducted at governmental schools in Gaza governorate.

### **Period of the study**

The study was commenced on December 2022, and completed in November 2024.

### **Inclusion criteria**

- Male and female English language teachers who are employed at governmental schools.
- Working in Gaza governorates.
- Have an experience in teaching English language.
- Willing to participate in the study and attend all the training program.

### **Tools of the study**

The researcher developed and used four tools for data collection:

Self-administered questionnaire: The questionnaire was used to measure the level of practicing authentic assessment strategies and tools among EL teachers.

Observation checklist: The researcher used observational checklist to assess the degree of implementing authentic assessment strategies and tools by the EL teachers in the classroom.

Rubrics: The researcher used rubrics tool to determine the changes in students' ability to speak English.

Personal Interview: The researcher used structured interview questions to gain in-depth information about the opinion of teachers regarding the effectiveness of authentic assessment.

### **Description of the questionnaire**

The questionnaire consisted of the following parts:

**First part:** Sociodemographic characteristics of the study participants (Gender, age, directorate, class, years of experience, previous training).

**Second part:** authentic assessment strategies. This part is subdivided to the following domains:

- Pencil and paper strategy: consists of (9) items.
- Performance-based assessment strategy: consists of (5) items.
- Observation strategy: consists of (4) items.
- Communication strategy: consists of (5) items.
- Self-assessment strategy: consists of (4) items.

**Third part:** Tools of authentic assessment: consists of (5) items.

### **Validity and reliability of the questionnaire**

The researcher examined face and content validity of the questionnaire by distributing the questionnaire to a group of experts in the area of research methodology and their comments were considered in finalizing the questionnaire.

The researcher performed a pilot study on 30 participants to examine the reliability of the questionnaire, and the results reflected that the questionnaire was reliable (Cronbach alpha = 0.845).

### **Study procedures:**

To be accomplished, the study passed through the following stages:

#### **Pre-test stage (Needs assessment):**

Before the training program, the researcher selected randomly 20 questionnaires for teachers who teach in west, east and north Gaza directorates to identify their awareness regarding the authentic assessment, in addition, observational checklist was used to assess the teachers' skills in using authentic assessment

in the classroom. Furthermore, observational checklist for 30 students was used to assess their competencies in speaking skills (pre-test).

#### **Intervention stage:**

In the light of the results of the pre-test the researcher designed a training program in authentic assessment. The training program included power-point presentations and practice sessions.

#### **Implementation of the training program:**

The researcher contacted the 20 teachers who was selected to participate in the training program, made an appointment and prepared the venue for training. The researcher explained the purpose of the study to the participants and asked them to sign the ethical approval. The training program was conducted in March and April 2023.

The training program included power point presentations, and practice on special skills of using the selected strategies and tools of authentic assessment which were included in the questionnaire.

#### **Post-test stage**

To evaluate the effectiveness of the training program, the researcher asked the participants to complete the questionnaires. Also, observation checklist was used to measure the participants' practices in the classroom concerning the assessment they used to assess their students. In Addition, rubrics was used to measure students' speaking skills after the training program. To have in-depth information about the study, interviews was conducted with five participants to explore their opinions about the training program, and how they evaluate themselves after the training.

#### **Stages of data collection**

The researcher collected data at two stages:

**Pre-test:** The study participants filled the questionnaire initially before implementation of the training program. The researcher also observed teachers to assess their classroom practices during English classes. In addition, rubrics was used to measure students' speaking skills before the training program.

**Post-test:** The participants filled the questionnaire after they have taken the training program. Also, the researcher conducted observation to notice the teachers' assessment of their student during English classes. In addition, rubrics was used to measure students' speaking skills after their teachers have taken the training program.

The researcher made sure that matching of pre-test and post-test was maintained to obtain accurate results. The researcher handled the questionnaires to the participated teachers, explaining the purpose of the study and providing clear instructions beforehand. Participants were given an estimated 15 minutes to complete each questionnaire.

#### **Data entry and statistical analysis**

The researcher used SPSS program (version 25) for data entry and statistical analysis. Before data analysis, the researcher performed data cleaning, coding and entry model.

**For quantitative part;** the researcher used descriptive statistics including frequencies, percentage, means, and standard deviation for socio-demographic variables.

Because the sample of the study is relatively small (20 teachers), the researcher used nonparametric tests including Wilcoxon test and Kruskal-Wallis test to compare the differences in results according to the studied variables.

Also, the researcher calculated the change percent to determine the degree of improvement and effectiveness of the training program.

**For qualitative part;** the researcher followed the following steps:

- **Data Preparation:** Reviewing collected data from interviews and organizing data systematically.
- **Coding and categorization:** Assigning codes to data that represent similar ideas or themes, and grouping related codes into categories to identify broader themes.

- **Thematic analysis:** Identifying recurring themes or patterns that provide meaningful insights. Analyzing frequency and occurrence of specific phrases, or ideas within textual data.
- **Triangulation of data:** comparing interview data with observational findings. It also provides a more holistic understanding of research problems, and identifies inconsistencies that lead to deeper insights.

By integrating different data types (quantitative and qualitative), the researcher gains a more nuanced, accurate, and comprehensive understanding of the phenomena under study.

#### **Ethical and administrative considerations**

The researcher obtained approval from the Islamic University of Gaza to carry out the study, then obtained approval from Ministry of Education. The participants have been informed that participation in the study is voluntary, and confidentiality of information is maintained. Consent form has been attached to each questionnaire.

## **Results**

### **Characteristics of EL teachers who participated in the training program**

The study included 20 EL teachers. Their characteristics are illustrated below.

**Table (1): Sociodemographic characteristics of study participants**

Variable	Number	Percentage (%)
<b>Gender</b>		
Male	7	35.0
Female	13	65.0
Total	20	100.0
<b>Age</b>		
25 years and less	3	15.0
26 – 35 years	6	30.0
36 – 45 years	11	55.0
Total	20	100.0
<b>Years of experience</b>		
5 years and less	3	15.0
6 – 10 years	5	25.0
11 – 15 years	12	60.0
Total	20	100.0
<b>Did you get any training on authentic assessment during your job as a teacher?</b>		
No	15	75.0
Yes	5	25.0
Total	20	100.0

Table (1) showed that 7 (35.0%) of study participants were male teachers and 13 (65.0%) were female teachers, 3 (15%) of study participants aged 25 years and less, 6 (30%) aged between 26 – 35 years, and 11 (55%) aged 36 - 45 years. Also, 12 (60%) of teachers have an experience between 11 – 15 years in English teaching, 5 (25%) have an experience between 6 – 10 years, and 3 (15%) have an experience of

5 years and less. Also, only 5 (25%) of teachers received previous training about authentic assessment while 15 (75%) of them did not receive training.

### Comparison of using authentic assessment strategies before and after the training program

The researcher used Wilcoxon test to compare the results before and after the training program, as demonstrated in the following tables.

**Table (2): Use of pencil and paper strategy before and after the training program (pre-post-test)**

No.	Items of pencil and paper strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Using tests and quizzes to assess students.	Pre	20	3.00	0.459	-3.800	0.000	33.3%
		Post	20	4.50	0.607			
2.	Using table of specifications when I construct tests.	Pre	20	2.20	0.410	-4.021	0.000	52.1%
		Post	20	4.60	0.503			
3.	Constructing tests according to learning objectives.	Pre	20	2.85	0.366	-3.841	0.000	35.2%
		Post	20	4.40	0.681			
4.	Constructing tests to assess listening skill for my students.	Pre	20	2.75	0.550	-4.030	0.000	40.8%
		Post	20	4.65	0.489			
5.	Constructing tests to assess speaking skill for my students.	Pre	20	3.95	0.510	-3.866	0.000	13.1%
		Post	20	4.55	0.605			
6.	Constructing tests to assess reading skill for my students.	Pre	20	3.00	0.459	-4.065	0.000	36.1%
		Post	20	4.70	0.571			
7.	Constructing tests to assess writing skill for my students.	Pre	20	3.10	0.553	-3.836	0.000	34.0%
		Post	20	4.70	0.571			
8.	Using essay questions when I construct tests.	Pre	20	2.65	0.587	-4.128	0.000	42.3%
		Post	20	4.60	0.681			
9.	Using objective questions when I construct tests.	Pre	20	2.50	0.607	-3.994	0.000	45.6%
		Post	20	4.60	0.598			
Total		Pre	20	2.77	0.222	-3.924	0.000	39.5%
		Post	20	4.58	0.383			

Wilcoxon test

Table (2) showed statistically significant improvement in using pencil and paper strategy after the training program. The highest improvement was in using table of specifications when constructing tests as mean score was 2.20 before the training program, increased to 4.60 after the training program with percentage change 52.1%, followed by improvement in using objective questions when constructing tests as mean score was 2.50 before the training program, increased to 4.60 after the training program with percentage change 45.6%. Overall, there was statistically significant improvement in using the pencil and paper strategy as mean score was 2.77 before the training program, increased to 4.58 after the training program with percentage change 39.5%.

**Table (3): Use of performance-based assessment strategy before and after the training program (pre-post-test)**

No.	Items of performance-based assessment strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Asking students to create presentations.	Pre	20	2.95	0.394	-3.981	0.000	36.5%
		Post	20	4.65	0.587			
2.	Asking students to do role play.	Pre	20	2.15	0.366	-4.028	0.000	52.2%
		Post	20	4.50	0.607			
3.	Asking students to simulate a real life situation using English language.	Pre	20	2.10	0.308	-3.993	0.000	53.8%
		Post	20	4.55	0.605			
4.	Asking students to demonstrate how they can do something.	Pre	20	2.85	0.489	-4.134	0.000	38.0%
		Post	20	4.60	0.598			
5.	Assessing students during their speech about something.	Pre	20	2.30	0.657	-4.005	0.000	50.5%
		Post	20	4.65	0.489			
Total		Pre	20	2.50	0.238	-3.935	0.000	45.5%
		Post	20	4.59	0.369			

Wilcoxon test

Table (3) indicated that the highest improvement was in asking students to simulate a real life situation using English language with mean score 2.10 before the training program, increased to 4.55 after the training program with percentage change 53.8%, followed by improvement in asking students to do role play with mean score 2.15 before the training program, increased to 4.50 after the training program with percentage change 52.2%. Overall, there was statistically significant improvement in using the performance-based assessment strategy as mean score was 2.50 before the training program, increased to 4.59 after the training program with percentage change 45.5%.

**Table (4): Use of observation strategy before and after the training program (pre-post-test)**

No.	Items of observation strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Observing students to get information about their use of language in the class.	Pre	20	2.10	0.308	-4.089	0.0001	55.3%
		Post	20	4.70	0.571			
2.	Determining a specific skill to be observed in each class.	Pre	20	2.35	0.489	-4.064	0.0001	48.9%
		Post	20	4.60	0.503			
3.	Preparing observation card to be used during observing students.	Pre	20	2.10	0.308	-4.042	0.0001	54.8%
		Post	20	4.65	0.489			
4.	Writing the observation notes in a record.	Pre	20	2.55	0.605	-3.954	0.0001	45.7%
		Post	20	4.70	0.470			
Total		Pre	20	2.27	0.279	-3.953	0.0001	51.2%
		Post	20	4.66	0.337			

Table (4) showed that the highest improvement was in observing students to get information about their use of language in the class with mean score 2.10 before the training program, increased to 4.70 after the training program with percentage change 55.3%, followed by improvement in preparing observation card

to be used during observing students with mean score 2.10 before the training program, increased to 4.65 after the training program with percentage change 54.8%. Overall, there was statistically significant improvement in using the observation strategy with mean score 2.27 before the training program, increased to 4.66 after the training program with percentage change 51.1%.

**Table (5): Use of communication strategy before and after the training program (pre-post-test)**

No.	Items of communication strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Interviewing students to assess their language.	Pre	20	3.15	0.366	-3.839	0.0001	30.7%
		Post	20	4.55	0.605			
2.	Using oral question to assess students' language skills.	Pre	20	3.10	0.447	-4.028	0.0001	34.7%
		Post	20	4.75	0.444			
3.	Using oral discussion to assess students' language skills.	Pre	20	2.50	0.607	-3.895	0.0001	45.6%
		Post	20	4.60	0.681			
4.	Asking probing questions to encourage students to think deeply and express themselves in English.	Pre	20	2.45	0.759	-3.882	0.0001	46.1%
		Post	20	4.55	0.759			
5.	Giving students enough time to think and express their answers.	Pre	20	2.50	0.688	-3.796	0.0001	43.8%
		Post	20	4.45	0.686			
Total		Pre	20	2.74	0.318	-3.935	0.0001	40.1%
		Post	20	4.58	0.498			

Table (5) showed that the highest improvement was in using oral question to assess students' language skills with mean score 3.10 before the training program, increased to 4.75 after the training program with percentage change 34.7%, followed by improvement in using oral discussion to assess students' language skills with mean score 2.50 before the training program, increased to 4.60 after the training program with percentage change 45.6%. Overall, there was statistically significant improvement in using the communication strategy as mean score was 2.74 before the training program, increased to 4.58 after the training program with percentage change 40.1%.

**Table (6): Use of self-assessment strategy before and after the training program (Pre-post-test)**

No.	Items of self-assessment strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Asking students to have portfolios.	Pre	20	2.60	0.589	-3.947	0.0001	43.4%
		Post	20	4.60	0.598			
2.	Asking students to write their strong and weak points about their learning.	Pre	20	2.90	0.447	-3.787	0.0001	33.3%
		Post	20	4.35	0.745			
3.	Asking students to correct their language mistakes themselves.	Pre	20	2.45	0.510	-3.921	0.0001	44.9%
		Post	20	4.45	0.759			
4.	Encouraging students to evaluate themselves and then give feedback.	Pre	20	2.85	0.489	-3.758	0.0001	32.9%
		Post	20	4.25	0.716			

Total	Pre	20	2.70	0.340	-3.937	0.0001	38.7%
	Post	20	4.41	0.575			

Table (6) showed that the highest improvement was in asking students to have portfolios with mean score 2.60 before the training program, increased to 4.60 after the training program with percentage change 43.4%, followed by improvement in asking students to correct their language mistakes themselves with mean score 2.45 before the training program, increased to 4.45 after the training program with percentage change 44.9%. Overall, there was statistically significant improvement in using the self-assessment strategy as mean score was 2.70 before the training program, increased to 4.41 after the training program with percentage change 38.7%.

**Table (7): Use of authentic assessment strategies before and after the training program (pre-post-test)**

No.	Strategy	n	Mean	SD	Z	P value	Percentage change	
1.	Pencil and paper strategy	Pre	20	2.77	0.222	-3.924	0.0001	39.5%
		Post	20	4.58	0.383			
2.	Performance-based assessment strategy	Pre	20	2.50	0.238	-3.935	0.0001	45.5%
		Post	20	4.59	0.369			
3.	Observation strategy	Pre	20	2.27	0.279	-3.953	0.0001	51.2%
		Post	20	4.66	0.337			
4.	Communication strategy	Pre	20	2.74	0.318	-3.935	0.0001	40.1%
		Post	20	4.58	0.498			
5.	Self- assessment strategy	Pre	20	2.70	0.340	-3.937	0.0001	38.7%
		Post	20	4.41	0.575			
<b>Total</b>		Pre	20	2.64	0.166	-3.922	0.0001	42.1%
		Post	20	4.56	0.389			

Table (7) illustrates that the greatest improvement was observed in the use of the observation strategy, with a mean score increasing from 2.27 before the training program to 4.66 after the program, reflecting a percentage change of 51.2%. This was followed by enhancements in the performance-based assessment strategy, where the mean score rose from 2.50 to 4.59, representing a 45.5% increase. The pencil-and-paper strategy also showed notable improvement, with the mean score climbing from 2.77 to 4.58, a percentage change of 39.5%. Similarly, the communication strategy saw a rise in mean score from 2.74 to 4.58, a 40.1% increase. The lowest improvement was recorded for the self-assessment strategy, which had a mean score of 2.70 before the training program and increased to 4.41 afterward, with a percentage change of 38.7%.

Overall, there was a statistically significant enhancement in the application of authentic assessment strategies, as the mean score improved from 2.64 before the training program to 4.56 after the program, marking a percentage change of 42.1%.

**Table (8): Scores for the tools of authentic assessment before and after the training program (pre-post-test)**

No.	Tools of authentic assessment	n	Mean	SD	Z	P value	Percentage change
1.	Pre	20	2.00	0.000	-4.093	0.000	56.9%

No.	Tools of authentic assessment		n	Mean	SD	Z	P value	Percentage change
	Using a checklist in assessing students.	Post	20	4.65	0.587			
2.	Using rubrics in assessing students.	Pre	20	2.00	0.000	-4.058	0.000	56.0%
		Post	20	4.55	0.686			
3.	Using rating scales in assessing students.	Pre	20	2.90	0.308	-3.987	0.000	35.5%
		Post	20	4.50	0.688			
4.	Using learning logs in assessing students.	Pre	20	3.25	0.550	-2.810	0.330	14.4%
		Post	20	3.80	0.523			
5.	Using anecdotal records in assessing students.	Pre	20	3.00	0.324	-3.787	0.000	32.5%
		Post	20	4.45	0.686			
Total		Pre	20	2.63	0.149	-3.936	0.000	40.0%
		Post	20	4.39	0.451			

Wilcoxon test

Table (8) showed that the highest improvement was in using checklist in assessing students with mean score 2.00 before the training program, increased to 4.65 after the training program with percentage change 56.9%, followed by improvement in using rubrics in assessing students with mean score 2.00 before the training program, increased to 4.55 after the training program with percentage change 56%. Overall, there was statistically significant improvement in using all the authentic assessment strategies as mean score was 2.63 before the training program, increased to 4.39 after the training program with percentage change 40%.

#### Authentic assessment strategies and selected sociodemographic variables

**Table (9): Differences in use of authentic assessment strategies related to gender (Mann-Whitney test)**

Strategy	Stage	Gender	n	m(SD)	P value
Pencil and paper strategy	Pretest	Male	7	2.76(0.162)	0.877
		Female	13	2.78(0.254)	
	Posttest	Male	7	4.20(0.422)	0.000*
		Female	13	4.79(0.109)	
Performance-based assessment strategy	Pretest	Male	7	2.51(0.279)	0.643
		Female	13	2.49(0.225)	
	Posttest	Male	7	4.25(0.320)	0.002*
		Female	13	4.76(0.256)	
Observation strategy	Pretest	Male	7	2.28(0.443)	0.393
		Female	13	2.26(0.160)	
	Posttest	Male	7	4.32(0.278)	0.001*
		Female	13	4.84(0.191)	
Communication strategy	Pretest	Male	7	2.65(0.250)	0.311
		Female	13	2.78(0.350)	
	Posttest	Male	7	4.05(0.472)	0.000*
		Female	13	4.86(0.189)	
Self- assessment strategy	Pretest	Male	7	2.50(0.381)	0.046*
		Female	13	2.80(0.272)	
	Posttest	Male	7	3.71(0.303)	0.000*

		Female	13	4.78(0.200)	
Total	Pretest	Male	7	2.58(0.137)	0.351
		Female	13	2.66(0.178)	
	Posttest	Male	7	4.10(0.256)	0.000*
		Female	13	4.81(0.116)	

Table (9) indicates a notable improvement in the use of the PP strategy following the training program, with female teachers demonstrating significantly better use of the PP strategy compared to male teachers ( $p = 0.000$ ). Similarly, there was a clear improvement in the use of the PBA strategy after the training, and female teachers showed significantly better use of the PBA strategy than male teachers ( $p = 0.002$ ). The use of the observation strategy also improved after the training program, with female teachers exhibiting significantly better implementation of the strategy compared to their male counterparts ( $p = 0.001$ ). Additionally, the use of the CS showed improvement after the training, and female teachers demonstrated significantly better use of the CS compared to male teachers ( $p = 0.000$ ).

Furthermore, there was a marked improvement in the use of the SA strategy, with female teachers expressing significantly better use of the strategy after the training program than male teachers ( $p = 0.000$ ).

Overall, all assessment strategies showed improvement after the training program, with female teachers demonstrating significantly better use of authentic assessment strategies compared to male teachers ( $p = 0.000$ ).

**Table (10): Differences in use of authentic assessment strategies related to experience (Kruskal-Wallis test)**

Strategy	Stage	Experience	n	M(SD)	Test statistics	P value
Pencil and paper strategy	Pretest	5years and less	3	2.81(0.256)	1.690	0.430
		6-10 years	5	2.64(0.121)		
		11-15 years	12	2.82(0.239)		
		Total	20	2.77(0.222)		
	Posttest	5years and less	3	4.00(0.587)	6.525	0.038*
		6-10 years	5	4.57(0.240)		
		11-15 years	12	4.74(0.223)		
		Total	20	4.58(0.383)		
Performance-based assessment strategy	Pretest	5years and less	3	2.46(0.416)	0.732	0.693
		6-10 years	5	2.56(0.167)		
		11-15 years	12	2.48(0.232)		
		Total	20	2.50(0.238)		

	Posttest	5years and less	3	4.06(0.305)	6.542	0.038*
		6-10 years	5	4.80(0.282)		
		11-15 years	12	4.63(0.305)		
		Total	20	4.59(0.369)		
Observation strategy	Pretest	5years and less	3	2.08(0.144)	4.904	0.086
		6-10 years	5	2.45(0.447)		
		11-15 years	12	2.25(0.184)		
		Total	20	2.27(0.279)		
	Posttest	5years and less	3	4.25(0.250)	4.899	0.086
		6-10 years	5	4.70(0.325)		
		11-15 years	12	4.75(0.301)		
		Total	20	4.66(0.337)		
Communication strategy	Pretest	5years and less	3	2.60(0.000)	1.555	0.459
		6-10 years	5	2.84(0.260)		
		11-15 years	12	2.73(0.374)		
		Total	20	2.74(0.318)		
	Posttest	5years and less	3	3.73(0.611)	9.014	0.011*
		6-10 years	5	4.52(0.268)		
		11-15 years	12	4.81(0.275)		
		Total	20	4.58(0.498)		
Self- assessment strategy	Pretest	5years and less	3	2.66(0.520)	0.602	0.740
		6-10 years	5	2.80(0.325)		
		11-15 years	12	2.66(0.325)		
		Total	20	2.70(0.340)		
	Posttest	5years and less	3	3.58(0.144)	5.427	0.066

		6-10 years	5	4.55(0.410)		
		11-15 years	12	4.56(0.534)		
		Total	20	4.41(0.575)		
Total	Pretest	5years and less	3	2.58(0.149)	0.582	0.748
		6-10 years	5	2.65(0.131)		
		11-15 years	12	2.64(0.190)		
		Total	20	2.64(0.166)		
	Posttest	5years and less	3	3.90(0.262)	6.505	0.039*
		6-10 years	5	4.61(0.265)		
		11-15 years	12	4.70(0.281)		
		Total	20	4.56(0.389)		

Table (10) showed that before the training program, teachers from different experiences showed low and moderate use of PP, PBA strategy, observation strategy, CS, SA strategy and the total score without statistical significant differences.

After the training program, there was a clear improvement in the use of PP strategy, and teachers with 11 – 15 years of experience showed significant better use of PP strategy ( $p= 0.038$ ), PBA strategy ( $p= 0.038$ ), CS ( $p= 0.011$ ), while improvement in use of observation strategy ( $p= 0.086$ ) and SA strategy ( $p= 0.066$ ) were not significant.

Generally, teachers with 11 – 15 years of experience expressed statistically significant improvement in using all the strategies after the training program compared to less experienced teachers ( $p= 0.039$ ).

#### Differences in using tools of authentic assessment related to sociodemographic factors

**Table (11): Differences in using tools of authentic assessment related to gender (Mann-Whitney test)**

Variable	Stage	Gender	n	m(SD)	P value
Tools of authentic assessment	Pretest	Male	7	2.60(0.115)	0.643
		Female	13	2.64(0.166)	
		Total	20		
	Posttest	Male	7	4.02(0.534)	0.030*
		Female	13	4.58(0.251)	
		Total	20		

Table (11) showed that: Before the training program, there were no statistical significant differences between male and female EL teachers in using tools of authentic assessment ( $p= 0.643$ ). Whereas, after the training program female teachers showed statistically significant higher use of tools of authentic assessment than male EL teachers ( $p= 0.030$ ).

**Table (12): Differences in using tools of authentic assessment related to experience (Kruskal-Wallis test)**

Variable	Stage	Experience	n	M(SD)	Test statistics	P value
Tools of authentic assessment	Pretest	5years and less	3	2.53(0.115)	3.975	0.137
		6-10 years	5	2.56(0.167)		
		11-15 years	12	2.68(0.133)		
		Total	20	2.63(0.149)		
	Posttest	5years and less	3	3.53(0.115)	7.762	0.021*
		6-10 years	5	4.56(0.328)		
		11-15 years	12	4.53(0.274)		
		Total	20	4.39(0.451)		

Table (12) showed that EL teachers from different experiences did not show statistically significant differences in using tools of authentic assessment before the training program ( $p= 0.137$ ), but after the training program, EL teachers who have 6 – 10 years of experience and those with 11 – 15 years of experience showed better use of tools of authentic assessment compared to teachers with an experience of 5 years and less ( $p= 0.021$ ).

### Observation of teachers

**Table (13): Observation of pencil and paper strategy (Pre-post-test)**

No.	Observed items of pencil and paper strategy		m(SD)	Z	P value
1.	Use written exercises to assess language skills.	Pre	0.45(0.510)	-2.828	0.005*
		Post	0.85(0.366)		
2.	Use written tests to assess speaking skills for students.	Pre	0.65(0.489)	-1.732	0.083
		Post	0.80(0.410)		
3.	Use essay questions in his/her written exercises, tests & quizzes.	Pre	0.30(0.470)	-3.000	0.003*
		Post	0.75(0.444)		
4.	Use objective questions in his/her written exercises, tests & quizzes.	Pre	0.65(0.489)	-1.732	0.083
		Post	0.80(0.410)		
Total		Pre	0.51(0.127)	-3.782	0.000*
		Post	0.80(0.130)		

Table (13) showed that there was statistically significant improvement in using written exercises to assess language skills ( $p= 0.005$ ), and Use essay questions in his/her written exercises, tests & quizzes ( $p= 0.003$ ), and the hole pencil and paper strategy after the training program ( $p= 0.000$ ).

**Table (14): Observation of performance-based assessment strategy (Pre-post-test)**

No.	Observed items of performance-based assessment		m(SD)	Z	P value
1.	Ask students to create presentations.	Pre	0.35(0.489)	-3.317	0.001*
		Post	0.90(0.308)		
2.	Ask students to do role play.	Pre	0.60(0.503)	-1.890	0.059
		Post	0.85(0.366)		
3.	Ask students to simulate a real life situation using English language.	Pre	0.55(0.510)	-2.530	0.011*
		Post	0.95(0.224)		
4.	Asking students to demonstrate how they can do something.	Pre	0.30(0.470)	-3.464	0.001*
		Post	0.90(0.308)		
5	Assess students during their speech about something.	Pre	0.50(0.513)	-3.162	0.002*
		Post	1.0(0.000)		
Total		Pre	0.46(0.234)	-3.852	0.000*
		Post	0.92(0.119)		

Table (14) showed that there was statistically significant improvement in using performance-based assessment strategy after the training program ( $p= 0.000$ ).

**Table (15): Observation of observation strategy (Pre-post-test)**

No.	Observed items of observation strategy		m(SD)	Z	P value
1.	Observe students to get information about their use of language in the class.	Pre	0.30(0.470)	-2.887	0.004*
		Post	0.80(0.410)		
2.	Determine a specific skill to be observed in each class.	Pre	0.45(0.510)	-2.646	0.008*
		Post	0.80(0.410)		
3.	Prepare observation card to be used during observing students.	Pre	0.35(0.489)	-2.449	0.014*
		Post	0.65(0.489)		
4.	Write the observation notes in a record.	Pre	0.35(0.489)	-2.646	0.008*
		Post	0.70(0.470)		
Total		Pre	0.35(0.125)	-3.905	0.000*
		Post	0.73(0.151)		

Table (15) showed that there was statistically significant improvement in using observation strategy after the training program ( $p= 0.000$ ).

**Table (16): Observation of communication strategy (Pre-post-test)**

No.	Observed items of communication strategy		m(SD)	Z	P value
1.	Interview students to assess their language.	Pre	0.35(0.489)	-3.317	0.001*
		Post	0.90(0.308)		
2.	Use oral question to assess students' language skills.	Pre	0.65(0.489)	-2.236	0.025*
		Post	0.90(0.308)		
3.	Use oral discussion to assess students' language skills.	Pre	0.10(0.308)	-3.464	0.001*
		Post	0.70(0.470)		

4.	Ask probing questions to encourage students to think deeply and express themselves in English.	Pre	0.45(0.510)	-3.162	0.002*
		Post	0.95(0.224)		
5.	Give students enough time to think and express their answers.	Pre	0.45(0.510)	-3.162	0.002*
		Post	0.95(0.224)		
Total		Pre	0.40(0.233)	-3.965	0.000*
		Post	0.88(0.119)		

Table (16) showed that there was statistically significant improvement in using communication strategy after the training program ( $p= 0.000$ ).

**Table (17): Observation of self-assessment strategy (Pre-post-test)**

No.	Observed items of self-assessment strategy		m(SD)	Z	P value
1.	Ask students to have portfolios.	Pre	0.35(0.489)	-2.877	0.004*
		Post	0.85(0.366)		
2.	Ask students to write their strong and weak points about their learning.	Pre	0.55(0.510)	-2.646	0.008*
		Post	0.90(0.308)		
3.	Ask students to correct their language mistakes themselves.	Pre	0.50(0.513)	-3.000	0.003*
		Post	0.95(0.224)		
4.	Encourage students to evaluate themselves.	Pre	0.50(0.513)	-2.828	0.005*
		Post	0.90(0.308)		
Total		Pre	0.47(0.255)	-3.671	0.000*
		Post	0.90(0.149)		

Table (17) showed that there was statistically significant improvement in using self-assessment strategy after the training program ( $p= 0.000$ ).

**Table (18): Observation of use of tools of authentic assessment (Pre-post-test)**

No.	Observed items of tools of authentic assessment		m(SD)	Z	P value
1.	Use a checklist in assessing students.	Pre	0.60(0.503)	-2.646	0.008*
		Post	0.95(0.224)		
2.	Use rubrics in assessing students.	Pre	0.50(0.513)	-2.646	0.008*
		Post	0.85(0.366)		
3.	Use rating scales in assessing students.	Pre	0.50(0.513)	-2.530	0.011*
		Post	0.90(0.308)		
4.	Use learning logs in assessing students.	Pre	0.60(0.503)	-2.121	0.034*
		Post	0.90(0.308)		
5.	Use anecdotal records in assessing students.	Pre	0.40(0.503)	-3.162	0.002*
		Post	0.90(0.308)		
Total		Pre	0.52(0.136)	-3.873	0.000*
		Post	0.90(0.121)		

Table (18) showed that there was statistically significant improvement in observed use of a checklist in assessing students ( $p= 0.008$ ), use rubrics in assessing students ( $p= 0.008$ ), use rating scales in assessing

students ( $p=0.011$ ), use learning logs in assessing students ( $p=0.034$ ), use anecdotal records in assessing students ( $p=0.002$ ), and use of all the tools of authentic assessment after the training program ( $p=0.000$ ).

### Observation of students

**Table (19): Summary of students' observation (Pre-post-test)**

No.	Observed items		n	Mean	SD	t	Percentage change	P value
1.	Using suitable vocabulary to express themselves in English language.	Posttest	30	2.53	0.507	5.757	20.9%	0.000
		Pretest	30	2.00	0.643			
2.	Use correct grammar when speak in English language in the class.	Posttest	30	2.80	0.407	13.573	42.8%	0.000
		Pretest	30	1.60	0.563			
3.	Using suitable expressions when speak in English language in the class.	Posttest	30	2.57	0.504	10.770	31.1%	0.000
		Pretest	30	1.77	0.679			
4.	Using the correct intonation when speak in English language in the class.	Posttest	30	2.60	0.498	7.992	35.7%	0.000
		Pretest	30	1.67	0.661			
5.	Using body language when speak in English.	Posttest	30	2.60	0.498	9.355	33.4%	0.000
		Pretest	30	1.73	0.740			
6.	Hesitation when speaking	Posttest	30	2.50	0.509	9.049	32.0%	0.000
		Pretest	30	1.70	0.837			
7.	Using eye contact during speaking English language.	Posttest	30	2.70	0.466	6.886	27.0%	0.000
		Pretest	30	1.97	0.765			
Total		Posttest	30	2.61	0.298	13.452	32.1%	0.000
		Pretest	30	1.77	0.548			

Table (19) presents a summary of the observation scores for students before and after the training program. The results show a statistically significant improvement in various aspects of English language use. Specifically, there was a 20.9% improvement ( $p=0.000$ ) in students' ability to use appropriate vocabulary to express themselves in English after the training. Additionally, there was a significant improvement of 42.8% ( $p=0.000$ ) in using correct grammar while speaking English in class.

The use of suitable expressions in English also improved significantly, with a percentage change of 31.1% ( $p=0.000$ ). Furthermore, students showed a 35.7% improvement ( $p=0.000$ ) in using correct intonation while speaking English in the classroom. Similarly, there was a 33.4% improvement ( $p=0.000$ ) in the use of body language when speaking English after the training program.

Students' hesitation while speaking decreased significantly, with a 47% improvement ( $p=0.000$ ). Additionally, there was a 27% improvement ( $p=0.000$ ) in using eye contact during English conversations after the training program.

Overall, there was a statistically significant improvement in the total score after the training program, with a percentage change of 32.1% ( $p=0.000$ ).

### Discussion

#### Use of authentic assessment strategies

The results indicated that teachers made significant improvements in using the observation strategy after the training program. The mean score increased from 2.27 before the training to 4.66 after, reflecting a 51.2% change. This finding aligns with Chen (2008), who stated that data collected through observation can provide valuable feedback, facilitating students' progress, promoting self-reflection, and enhancing

active involvement in their own learning. Interviewed teachers confirmed that observation is crucial for all educators, as it provides comprehensive insights into students' abilities and skills, leading to more accurate assessments.

Additionally, the results showed that teachers significantly increased their use of the PP strategy after the training, with the mean score rising from 2.77 to 4.58, a 39.5% improvement. This outcome is consistent with Hussain et al. (2019), who reported that paper-and-pencil assessments were underused by teachers. The interviews revealed that while teachers initially used the PP strategy, they recognized the need for improvement after the training.

Similarly, there was a notable increase in the use of PBA strategy, with the mean score rising from 2.50 to 4.59, a 45.5% improvement. This result aligns with Lund & Kirk (2019), who found that PBA allows students to showcase their skills while providing educators with valuable feedback. Interviews revealed that teachers were not previously using PBA effectively, often relying on written tests or informal observations. However, after the training, most teachers recognized the value of PBA in providing meaningful feedback on their teaching strategies.

The CS was also used to a greater extent following the training, with the mean score increasing from 2.74 to 4.58, a 40.1% improvement. In comparison to other studies, Zaim et al. (2020) noted that teachers face challenges in assessing communicative functions, which aligns with the interviewed teachers' previous reliance on written tests for assessing speaking skills. After the training, teachers acknowledged that the communicative strategy was more effective for evaluating speaking proficiency.

Moreover, the use of SA strategies also increased significantly, with the mean score rising from 2.70 to 4.41, a 38.7% improvement. Most teachers, except one, agreed on the importance of self-assessment for evaluating students' abilities. However, a few teachers expressed concerns that some students might exaggerate or misrepresent their abilities during self-assessment.

In summary, the teachers demonstrated a significant increase in their use of all authentic assessment strategies after the training program, with an overall improvement of 42.1%. The interviews with teachers confirmed these results, with teachers expressing the value of the training program and the practical application of authentic assessment strategies in their classrooms. This outcome is consistent with findings by Taufina & Chandra (2017), who reported a similar increase in teachers' proficiency with authentic assessment after training. Additionally, Asante (2023) found that most teachers were familiar with authentic assessment and applied it effectively, encouraging student collaboration and real-world applications. Similarly, Prawisanthi et al. (2021) and Maulidhawati et al. (2021) noted that while teachers implemented authentic assessments, they still relied on traditional methods for evaluating language proficiency. However, studies by Al-Shatnawi & Bani Khalaf (2018) highlighted the positive impact of training programs on teachers' use of authentic assessment strategies.

#### **Association between authentic assessment strategies and sociodemographic variables**

Female EL teachers expressed significant better use of authentic assessment strategies. Also, EL teachers with longer years of experience showed significant better use of authentic assessment strategies, while age and previous training did not make difference in using the strategies. This result was consistent with the results of Al-Shatnawi & Bani Khalaf (2018) who found that female teachers and those who received previous training obtained higher scores on authentic assessment, while experience did not make significant difference in use of authentic assessment. Whereas, the results of Asante (2023) indicated that there was no statistically significant difference in teachers' knowledge of authentic assessment related to gender.

#### **Tools of authentic assessment**

The results demonstrated a significant improvement in the use of authentic assessment tools, with the total mean score rising from 2.63 (52.6%) before the training program to 4.39 (87.8%) afterward,

reflecting a 40% increase. This aligns with the findings of Taufina & Chandra (2017), which showed a similar improvement in teachers' assessment scores, from an average of 39.06 before the mentoring program to 64.78 after it. All interviewed teachers reported that they were unfamiliar with authentic assessment tools prior to the training but began applying them in their student assessments once they were introduced.

Additionally, female EL teachers and those with more years of experience showed significantly better use of authentic assessment tools. However, age and previous training did not appear to influence their use of these tools.

Irwana (2006) identified four primary forms of assessment in language learning: performance assessment, product assessment, portfolio assessment, and assessment tests. A common challenge teachers face in implementing authentic assessment is their limited experience, as well as the considerable time and effort needed to manage student grades while minimizing subjectivity in final evaluations. To combat subjectivity, Sahyoni & Zaim (2017) suggested using rubric models that clearly define achievement levels, allowing teachers to assess students' proficiency more accurately based on specific criteria.

Teachers often struggle to use a variety of assessment tools and tend to apply the same rubric to both speaking and writing tasks. However, using a single rubric for different types of assessments can limit the accuracy of evaluating students' skills, as each task type requires distinct achievement indicators (Paramartha & Pratiwi, 2017).

#### **Observation of teachers**

Observations of teachers revealed a significant increase in the use of authentic assessment strategies after the training program, with the improvement being statistically significant. These results are consistent with the findings of Maulidhawati et al. (2021), who noted that all 8th-grade teachers understood the concept of authentic assessment and regularly employed methods such as storytelling and direct oral interviews in English classes. Similarly, Prawisanthi et al. (2021) found that EL teachers felt confident in their understanding of authentic assessment, with an implementation rate of 78.52%.

In contrast, Juaidy and Afona (2020) reported that teachers' experience with authentic assessment was moderate, and there were no significant differences in experience levels based on sociodemographic factors, except for years of experience. A comparison of assessment practices before and after the training program showed significant improvements across all components of the authentic assessment strategies, reflecting a greater recognition of the importance and effectiveness of these strategies.

These findings highlight the value of authentic assessment, which integrates both process and product assessment within thematic learning. While teachers may face challenges in applying authentic assessment, comprehensive training can help address these obstacles. Regular, ongoing training programs focused on designing effective assessment tools are essential for preparing teachers to accurately evaluate students' knowledge and skills.

#### **Observation of students:**

Student observations indicated a significant improvement in English-speaking skills following the training program for teachers, with a 32.1% increase. Authentic assessment provides numerous benefits by fostering creative and critical thinking through engagement with complex, real-world tasks. It allows students to apply a variety of skills and competencies within assessment activities, enhancing their self-efficacy and work-readiness through practical experiences (The University of Melbourne, 2022). These findings align with Maulidhawati et al. (2021), who emphasized the effectiveness of authentic assessment in capturing students' English proficiency, particularly in productive skills.

Several studies (Inayah et al., 2019; Elsara & Maiwen, 2019; Afriadi et al., 2021) have demonstrated the high effectiveness of authentic assessment in evaluating speaking abilities and recommend its application

for assessing other language skills. Authentic assessment nurtures practical problem-solving, communication, critical thinking, and collaboration skills. From a socio-cultural perspective, it promotes peer learning, social engagement, and deeper understanding through active participation (Bohemia & Davison, 2012).

Sukma et al. (2022) found that students responded positively to the authentic assessment model, viewing it as effective and engaging, particularly for evaluating reading comprehension skills in junior high school. Speaking is a vital skill for English learners, but assessing it presents unique challenges due to the various factors influencing foreign language speaking abilities. Teachers must design and implement appropriate authentic assessments tailored to language learning (Zaim et al., 2020). Authentic assessments also provide opportunities for students to construct their own knowledge and enhance their understanding of concepts (Azim & Khan, 2012). This approach encourages active learning, with teachers taking on a facilitating role rather than being the central figure in instruction. Through authentic assessment, students are more engaged in the learning process, especially in collaborative activities (Kearney & Perkins, 2010).

Rukmini & Saputri (2017) highlighted that teachers often use authentic assessments for speaking skills, incorporating tasks like role plays and debates to directly observe students' language production. They argued that speaking assessments should focus on students' ability to interpret and convey meaning in interactive settings, making authentic assessment an invaluable tool for developing speaking skills.

In summary, assessment provides significant benefits for both teachers and students. For teachers, it enhances their ability to track student progress, offers valuable insights for refining teaching methods, and helps in selecting the most effective strategies for different student groups. For students, feedback from assessments improves their understanding and knowledge, fosters self-evaluation, and helps them view assessment as an integral part of the learning journey, guiding their approach to acquiring new skills and knowledge (Thomas et al., 2004).

The study concluded that teachers' use of authentic assessment strategies had improved, benefitting both educators and students. It recommended expanding the training program to other areas and frequently applying authentic assessment to maximize its impact.

## References

Afriadi, Z., Arifuddin, A., & Nuriadi, N. (2021, May). The implementation of authentic assessment in speaking skills at MTS As-Suyuthy Ireng. In *2nd Annual Conference on Education and Social Science (ACCESS 2020)* (pp. 241-244). Atlantis Press.

Afshar, H. S., & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, 70, 101042.

Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications*, 11(4).

AlKhateeb, M. A. (2018). The effect of using performance-based assessment strategies to tenth-grade students' achievement and self-efficacy in Jordan. *Kıbrıslı Eğitim Bilimleri Dergisi*, 13(4), 489-500.

Al-Najjar, A. (2018). *The Role of Teacher Questioning in Enhancing Students' Critical Thinking Skills in EFL Classrooms*. *Journal of Educational and Psychological Studies*, 12(4), 547-563.

Alonso-Tapia, J., & Panadero, E. (2010). Effects of self-assessment scripts on self-regulation and learning. *Infancia y aprendizaje*, 33(3), 385-397.

Al-Zoubi, M. S. (2019). Classroom authentic assessment strategies and tools used by english language teachers in Jordan. *International Journal of Language and Linguistics*, 6(4), 34-43.

- Arter, J. A., & Chappuis, J. (2006). *Creating & Recognizing Quality Rubrics*. Assessment Training Institute, Inc Series.
- Asante, G. (2023). Knowledge and practices of authentic assessment in junior high schools. *Open Access Library Journal*, 10(5), 1-14.
- Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, 110, 154-169.
- Azim, S., & Khan, M. (2012). Authentic assessment: An instructional tool to enhance students learning. *Academic Research International*, 2(3), 314.
- Aziz, M. N. A., Yusoff, N. M., & Yaakob, M. F. M. (2020). Challenges in Using Authentic Assessment in 21st Century ESL Classrooms. *International Journal of Evaluation and Research in Education*, 9(3), 759-768.
- Bailey, K. M. (1998). Learning about language assessment: Dilemmas, decisions, and directions. *Heinle & Heinle Publishers google scholar*, 2, 653-675.
- Baker, L., Dreher, M. J., & Guthrie, J. T. (Eds.). (2000). *Engaging young readers: Promoting achievement and motivation*. Guilford Press.
- Balgan, A., Renchin, T., & Ojgoosh, K. (2022). An experiment in applying differentiated instruction in STEAM disciplines. *Eurasian Journal of Educational Research*, 98(98), 21-37.
- Barnawi, O. Z. (2022). Branding in Transnational English Medium Instruction-Oriented Universities in the Arabian Gulf: Implications for Language Policy. *Eurasian Journal of Applied Linguistics*, 8(1), 58-72.
- Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American journal of psychology*, 128(2), 241-252.
- Bohemia, E., & Davison, G. (2012). Authentic Learning: The gift project. *Design and Technology Education*, 17(2), 49-61.
- Brown, H. D., & Abeywickrama, P. (2004). Language assessment. *Principles and Classroom Practices*. White Plains, NY: Pearson Education, 20.
- Brown, J. D. (1998). *New Ways of Classroom Assessment*. *New Ways in TESOL Series II. Innovative Classroom Techniques*. TESOL, 1600 Cameron Street, Suite 300, Alexandria, VA 22314-2751.
- Butakor, P. K., & Caesar, J. (2021). Analysing Ghanaian teachers' perceived effects of authentic assessment on student performance in Tema Metropolis: Analysing Ghanaian teachers' perceived effects of authentic assessment. *International Journal of Curriculum and Instruction*, 13(3), 1946-1966.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325.
- Chen, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language teaching research*, 12(2), 235-262.
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360.
- Ekawati, D. (2017, April). The Implementation of authentic assessment in vocational high school 1 Kuala Cenaku. In *English Language and Literature International Conference (ELLiC) Proceedings* (Vol. 1, pp. 82-86).
- Elsara, W., & Maiwen, S. (2019, March). English Teachers' Difficulties in Implementing Authentic Assessment in Speaking Skill. In *Seventh International Conference on Languages and Arts (ICLA 2018)* (pp. 255-259). Atlantis Press.
- Falsgraf, C. (2005). *Why a national assessment summit? New visions in action*. *National Assessment Summit. Meeting conducted in Alexandria, Va.*

Furtak, E. M., Morrison, D., & Iverson, H. (2013, April). Challenges in developing classroom assessments linked to multidimensional learning progressions. In *National Association of Research on Science Teaching Annual International Conference, Puerto Rico*.

Gallardo, K. (2020). Competency-based assessment and the use of performance-based evaluation rubrics in higher education: Challenges towards the next decade. *Problems of Education in the 21st Century*, 78(1), 61-79.

Gallavan, N. P. (Ed.). (2008). *Developing performance-based assessments, grades 6-12*. Corwin Press.

Gezie, A., Khaja, K., Chang, V. N., Adamek, M. E., & Johnsen, M. B. (2012). Rubrics as a tool for learning and assessment: What do Baccalaureate students think?. *Journal of Teaching in Social Work*, 32(4), 421-437.

Gorman, G.E., & Clayton, P. (2007). *Literary and Linguistic Computing*, 22(2), 246–248.

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, 23, 531-557.

Hedge, T. (2000). *Teaching and Learning in the Language Classroom*. Oxford University Press.

Herlo, D. D. (2014). Paragogy. A new theory in educational sciences. *Educația Plus*, 10(1), 35-41.

Herrera, S. G., Murry, K. G., & Cabral, R. M. (2013). *Assessment Accommodations for Classroom Teachers of Culturally and Linguistically Diverse Students*. Pearson Higher Ed.

[https://www.queensu.ca/teachingandlearning/modules/assessments/25\\_s3\\_03\\_self\\_assessment.htm](https://www.queensu.ca/teachingandlearning/modules/assessments/25_s3_03_self_assessment.htm)

l. Accessed on 12.8.2023.

Hussain, S., Shaheen, N., Ahmad, N., & Islam, S. U. (2019). Teachers' classroom assessment practices: challenges and opportunities to classroom teachers in Pakistan. *Dialogue*, 14(1), 88-97.

Idham, F. Islami, Nadrun, Darmawan (2015). The use of authentic assessment in English writing skill to the eleventh grade students. *E-Journal of English Language Teaching Society*, 3(1), 1-13.

Inayah, N., Komariah, E., & Nasir, A. (2019). The practice of authentic assessment in an EFL speaking classroom. *Studies in English Language and Education*, 6(1), 152–162.

Indiana University Bloomington, (2023). Center for Innovative Teaching and Learning. Available from: <https://citl.indiana.edu/teaching-resources/assessing-student-learning/authentic-assessment/index.html>.

Irwana, E. (2006). Penerapan asesmen otentik dalam pembelajaran bahasa Indonesia; Studi kasus di SMPN 1 Wlingi, Kecamatan Wlingi, Kabupaten Blitar. Skripsi. Fakultas Sastra, Universitas Negeri Malang.

Jafarpour, A. (2003). Is the test constructor a facet? *Language Testing*, 20(1), 57-87.

Jaidey, M., & Afona, S. (2020). Degree of principals and teachers' assessment of the experience of authentic assessment at schools in Qalqilia's governorate and ways to develop it. Al Azhar University, Gaza, *Journal of Human Sciences*, 22(1), 1-34.

Jönsson, A., & Panadero, E. (2017). The use and design of rubrics to support assessment for learning. *Scaling up assessment for learning in higher education*, 99-111.

Kabir, M. H. (2012). Necessity of initiating rating scale for more reliable assessment of writing skill at HSC level: a case study. *Liuc Studies*, (6), 35-52. <https://doi.org/10.3329/iiucs.v6i0.12247>.

Katal, A., Singh, V. K., Choudhury, T., & Imran, F. (2022). Enhancing Teaching and Learning through Peer Observation: An Indian Case Study. *Education Research International*, 2022:7825178.

Kearney, S., & Perkins, T. (2010). Improving engagement: the use of authentic self and peer assessment for learning to enhance the student learning experience. *Study on Primary Education*, 38(7), 1-11.

- Kniep, J. I. S. K. A., & Janssen, T. A. N. J. A. (2014). Effects of observational learning on students' use of and attitude towards reading and learning strategies. *L1-Educational Studies in Language and Literature, 14*, 1-21.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: what should they look like and where should the criteria come from? *Assessing Writing, 16*(2), 81-96.
- Koh, K. H. (2017). Authentic assessment. In *Oxford research encyclopedia of education*. <https://oxfordre.com/education/view/10.1093/acrefore/97801>.
- Koné, K. (2021). Exploring the impact of performance-based assessment on Malian EFL learners' motivation. *Advances in Language and Literary Studies, 12*(3), 51-64.
- Larkin, T.L. (2014). The student conference: A model of authentic assessment. *International Journal of Engineering Pedagogy, 4*(2), 36-46.
- Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing, 26*, 67-79.
- Lund, J. L., & Kirk, M. F. (2019). *Performance-based assessment for middle and high school physical education*. Human Kinetics Publishers.
- McDonald, K., & Park, S. Y. (2016). Empowering higher education foreign language teachers with effective peer observation and feedback practice. Defense Language Institute Foreign Language Center. Manchester Metropolitan University, (2023). What is authentic assessment?. Available from: <https://www.mmu.ac.uk/about-us/professional-services/uta/assessment/authentic>, Accessed 19.7.2023.
- Maulidhawati, R., Prastikawati, E. F., & Budiman, T. C. S. (2021). Authentic assessments in English language teaching: A case in smp negeri 6 semarang. *Journal of English Education and Linguistics, 2*(1), 67-78.
- McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (Vol. 2). John Wiley & Sons.
- Mertler, C. A. (2002). Classroom assessment literacy inventory. (Adapted from the Teacher Assessment Literacy Questionnaire (1993), by Barbara S. Plake & James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education & the W.K. Kellogg Foundation.
- Mertler, C. A. (2003). Pre-service versus in-service teachers' Assessment literacy: Does classroom experience make a Difference? *Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH*.
- Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory. *Paper presented at the annual meeting of the American Educational Research Association*, April, Montreal, QC, Canada.
- Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centered framework for feedback literacy. *Assessment & Evaluation in Higher Education, 45*(4), 527-540.
- Muis, K. R., Ranellucci, J., Trevors, G., & Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction, 38*, 1-13.
- Navidinia, H., Mobaraki, M., & Malekzadeh, F. (2019). Investigating the effect of noticing on EFL students' speaking accuracy. *International Journal of Instruction, 12*(1), 83-98.
- Nguyen, T. T. K., & Phan, H. M. (2020). Authentic assessment: A real life approach to writing skill development. *International Journal of Applied Research in Social Sciences, 2*(1), 20-30.
- Nurgiantoro, B. & Suyata, P. (2009). Pengembangan model asesmen otentik dalam pembelajaran bahasa. *Cakrawala Pendidikan, 28*(3), 224-237.

- Ojung'a, J., & Allida, D. (2017). A survey of authentic assessments used to evaluate English language learning in Nandi central sub-county secondary schools. *Kenya: Baraton Interdisciplinary Research Journal*, 7, 1-11.
- Ortega, D. P., & Minchala, O. E. (2017). Assessing students in an authentic and ongoing manner in the English classroom. *Theory and Practice in Language Studies*, 7(3), 159-165.
- Paramartha, A. A. G. Y. & Pratiwi, N. P. A. (2017). Problematika penerapan asesmen dalam Kurikulum 2013 oleh guru bahasa Inggris di SMPN 2 Singaraja. *Prosiding Seminar Nasional Riset Inovatif*, 2017, pp. 746-751.
- Parmelee, D., Michaelsen, L. K., Cook, S., & Hudes, P. D. (2012). Team-based learning: A practical guide: AMEE Guide No. 65. *Medical Teacher*, 34(5), e275-e287.
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. Building cathedrals: Compassion for the 21st century. Available from: <http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html>.
- Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63, 84-85.
- Porcello, D., & Hsi, S. (2013). Crowdsourcing and curating online education resources. *Science*, 341(6143), 240-241.
- Prawisanthi, N., Dewi, N., Paramartha, A., & Wahyuni, L. (2021). The implementation of English teachers' authentic assessment in junior high schools. *International Journal of Language and Literature*, 5(4), 212-226.
- Puengpipattrakul, W. (2019). *The influence of review strategies on students' English-language writing performance in a first-year university undergraduate Thai EFL writing context* (Doctoral dissertation, University of Birmingham).
- Rawekar, A., Garg, V., Jagzape, A., Despande, V., Tankhiwale, S., & Chalak, S. (2013). Team based learning: A controlled trial of active learning in large group settings. *Journal of Dental and Medical Sciences*, 7(4), 42-48.
- Reynisdóttir, B. B. (2016). *The efficacy of authentic assessment: a practical approach to second language testing* (Doctoral dissertation).
- Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American educator*, 36(1), 12.
- Rukmini, D., & Saputri, L. (2017). The authentic assessment to measure students' English Productive skills based on 2013 curriculum. *Indonesian Journal of Applied Linguistics*, 7(2), 263-273.
- Rusilowati, Ani. (2013). Masih Banyak Guru Kesulitan Implementasi Kurikulum Baru. Okezone. Diperoleh 21 Januari 2016, dari <http://www.okezone.com>.
- Sachs, J., & Parsell, M. (Eds.). (2013). *Peer review of learning and teaching in higher education: International perspectives* (Vol. 9). Springer Science & Business Media.
- Sahyoni, S., & Zaim, M. (2017). Authentic assessment of speaking skills for grade VII/MTsN. *Komposisi Journal Pendidikan Bahasa Sastra dan Seni*, 18(1), 15.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Australia Language Testing* 30(3) 309–327.
- Setiawan, D., & Hadikusuma, Z. (2015). Pengembangan Penilaian Autentik Berbasis Kurikulum 2013 di Kelas IV SD Kota Pekanbaru. *Jurnal Tematik*, Tahun April 2015. Medan: PPs Pendidikan Dasar Unimed.
- Shing, Y. (2011). *Observation in Language Classrooms: A Guide to Effective Language Teaching and Learning*. TESOL Journal.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371–391.

- Stanley, T. (2021). *Performance-based assessment for 21st-century skills*. Routledge.
- Stephens, N.M. (2021). Paper- and-pencil testing: still around?. Assessment Systems Corporation. Available at: <https://assess.com/paper-and-pencil-testing>.
- Stiggins, R., & Conklin, N. (2019). In teachers' hands: Investigating the practices of classroom assessment. *Albany: State University of New York Press*.
- Sukma, Abbas, A., Nurhayati, K., & Gheisari, A. (2022). Development of authentic assessment in local wisdom-based reading learning. *Education Research International*, 2022, 9405583.
- Sundberg, M. D. (2002). Assessing student learning. *Cell Biology Education*, 1, 11–15.
- Sutadji, E., Susilo, H., Wibawa, A.P., Jabari, N.A.M., & Rohmad, S.N. (2021). Adaptation strategy of authentic assessment in online learning during the covid-19 pandemic. *Journal of Physics: Conference Series*, 1810(1), ID 012059.
- Taufina, T., & Chandra, C. (2017). The implication of authentic assessment in thematic integrated learning process at lower level elementary school. *Advances in Social Science, Education and Humanities Research (ASSEHR)*, volume 169. In *International Conference of Early Childhood Education (ICECE 2017)* (pp. 140-143). Atlantis Press.
- Teaching Today. (2011). Student Learning Log. [http://www.glencoe.com/sec/teachingtoday/downloads/pdf/learning\\_log.pdf](http://www.glencoe.com/sec/teachingtoday/downloads/pdf/learning_log.pdf)
- The University of Melbourne, (2022). Authentic assessment. Available from: <https://le.unimelb.edu.au/learning-teaching-assessment/assessment-and-feedback/authentic-assessment>. Accessed 19.7.2023.
- Thomas, J., Allman, C., & Beech, M. (2004). *Assessment for the diverse classroom: A handbook for teachers*. Tallahassee, FL: Florida Department of Education, Bureau of Exceptional Education and Student Services. Retrieved from [http://www.fldoe.org/ese/pdf/assess\\_diverse.pdf](http://www.fldoe.org/ese/pdf/assess_diverse.pdf)
- Tomlinson, B. (2011). *Material development in language teaching*. 2<sup>nd</sup> edition, Cambridge University Press. Available from: [www.cambridge.org/elt/cltl](http://www.cambridge.org/elt/cltl).
- University of New South Wales, (2023). *Assessing authentically*. Available from: <https://www.teaching.unsw.edu.au/authentic-assessment>.
- Wahyuni, Y. (2019). *An analysis on teachers authentic assessments in English teaching at SMP It Khairunnas in academic year 2018/2019 (Doctoral dissertation, Bahasa Inggris)*.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>.
- William, D. (2016). What are the advantages of authentic assessment over standardized testing? Retrieved from [http://oureverydaylife.com/advantages\\_authentic\\_assessment\\_over\\_standardized\\_testing\\_20187.html](http://oureverydaylife.com/advantages_authentic_assessment_over_standardized_testing_20187.html).
- Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *Journal of Effective Teaching*, 7(1), 3-14.
- Wu, X. M., Dixon, H. R., & Zhang, L. J. (2021). Sustainable development of students' learning capabilities: the case of university students' attitudes towards teachers, peers, and themselves as oral feedback sources in learning English. *Sustainability*, 13(9), 5211.
- Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia-Social and Behavioral Sciences*, 116, 2998-3004.
- Yiend, J., Weller, S., & Kinchin, I. (2014). Peer observation of teaching: The interaction between peer review and developmental models of practice. *Journal of Further and Higher Education*, 38(4), 465-484.
- Zaim, M. & Refnaldi. (2016). Teachers' need on authentic assessment for speaking skills. *Proceeding International Seminar on Languages and Arts (ISLA) 5*. Padang: FBS UNP Press.

Zaim, M. (2017). Implementing scientific approach to teach English at senior high school in Indonesia. *Asi. Soc Sci*, 13(2), 33-40.

Zaim, M., Refnaldi, & Arsyad, S. (2020). Authentic Assessment for Speaking Skills: Problem and Solution for English Secondary School Teachers in Indonesia. *International Journal of Instruction*, 13(3), 587-604.

Zhang, D., & Zhang, L. J. (2019). Metacognition and self-regulated learning (SRL) in second/foreign language teaching. *Second handbook of English language teaching*, 883-897.

Zhang, L. J., & Zhang, D. (2018). *Metacognition in TESOL: Theory and practice*. In Liontas JI, & Shehadeh A.(Eds.), *The TESOL encyclopedia of English language teaching*, Vol. II: Approaches and methods in English for speakers of other languages. Hoboken.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1), 166-183.

الشطناوي، شيرين يوسف خالد وبني خلف، محمود حسن مصطفى (2018). Al-Shatnawi & Bani Khalaf. تصميم برنامج

تدريبي لمعلمي العلوم قائم على استراتيجيات التقويم الواقعي وقياس أثره على كفاءتهم الذاتية وممارستهم التقييمية، (رسالة دكتوراه غير منشورة) جامعة اليرموك، إربد، الأردن.