

Boosting Language Assessment Literacy and Positive Backwash: A Skill-Based Language Teaching Approach for EFL Graduate Students

Dr. Dalia Ali Maher Abbass Mohammed

Associate professor of Curriculum & EFL Instruction – Faculty of Education – Mina University- Egypt

ABSTRACT

The present study aimed at investigating the effectiveness of using the skill-based language teaching approach to boost EFL graduate students' language assessment literacy and positive backwash effect. The study utilized the quasi-experimental research method (pretest-posttest non-treatment group design). Thirty participants enrolled in the professional diploma at the Faculty of Education, Minia University were randomly assigned to two intact groups: a treatment group (n=15) and a non-treatment group (n=15). The participants in the treatment group were trained and instructed using a skill-based language teaching program designed by the researcher whereas their counterparts in the non-treatment group did not receive such training as they received regular instruction. Instruments of the study included a needs-analysis questionnaire for assessment literacy, a test of language assessment literacy, and a scale of backwash effect. The findings revealed that the participants in the treatment group significantly surpassed their counterparts in the non-treatment group in the post-performance of the test of language assessment literacy, and the scale of backwash effect. Suggestions for further research and recommendations were also presented.

Keywords: *skill-based language teaching approach, language assessment literacy, backwash effect*

Introduction

Assessment plays a crucial role in education, politics, economics, and all fields of society. More specifically, assessment and its outcomes have a very significant role in the teaching and learning processes. It involves collecting information about what learners know and can do before, during, or after instruction, and passing judgement for the sake of making decisions. Continuous learning is always supported by assessment tools. While formative assessments provide ongoing feedback to guide learning along the way, summative assessments measure learners' mastery at the end of specific learning segments.

Formative assessment most often takes place inside the classroom as it focuses on the actual learning process while summative assessment tends to happen at the end of a learning cycle to measure learners' learning outcomes. Formative and summative assessments, like foreign language tests, can be powerful drivers of educational practice. They can affect teachers' methodologies, prompt syllabus revisions, and motivate teachers, administrators, students, and parents alike. However, this influence can lead to both positive pressure for improvement and potential stress for all stakeholders involved.

Assessment, in all fields, covers the activities of data collection and decision making. There are traditional (formal) assessment methods which are contrasted to authentic (informal/performance-based/ alternative) assessment methods. Traditional assessment methods have emphasized mainly the cognitive domain evaluation mode while authentic assessment addresses more holistic domains of cognitive – affective – psychomotor evaluation method. Thus, authentic assessment comprehensively

assesses learning input, process, and output considering learners characteristics and level continuously and sustainably while teaching and learning process. Then, teachers need to use methods other than tests for a better insight and right decision (e.g. mini-projects, reports, research papers, products, portfolios, etc.) (Suwartono & Riyani, 2019:113).

Language Assessment

Assessment is sometimes misunderstood with testing. Tests occur at definite times during instruction and learners are expected to show their best responses for gaining high scores and evaluation. Assessment, on the other hand, is a continuous incidental or intended process encompassing a wide range of learners' responses, i.e. answers to questions, comments on ideas presented, oral discussion, written responses, etc. (Şenel & Tütüniş: 2011:46). Regardless of any type of assessment adopted, teachers of English have been supposed to measure both competence as well as performance in a communicative way that might encounter learners in their own real-life situation. The emphasis is not only on linguistic accuracy, but on the ability to function effectively through language in particular contexts of situation.

Language assessment is affected by the social, cultural, and educational contexts surrounding tests. However, research on language testing for quite a long time tackled only the measurement issues while neglecting the various aspects that language tests can play in society. Language testing has focused on the analysis of a particular group of test-takers and the way of incorporating such information in designing language exams. The content of language testing determines the choice of language teaching content to a great extent.

Language exam scores are affected by multiple factors. Dawadi (2021) and Cheng (2000, as cited in Alqahtani, 2021:23) have pointed out that these factors include the exam's features and contexts, the test-takers' characteristics, the tactics used to complete the exams, and the inferences that instructors wish to draw from them as well as the interrelationships between these factors. Thus, language exam scores alone do not provide a complete picture of a student's language ability. Even in the modern age, tests influence education and employment. Despite all criticisms leveled against them, exams continue to dominate the educational sector in most nations. Given the critical decisions attached to exams, it is evident that tests control many facets of life.

Both the development of language teaching and language testing have undergone three stages marked by different teaching and testing systems. The communicative language teaching approach (CLT), emerged at the end of the 20th century, directs the current language teaching and testing systems towards competence. This means the transformation from language knowledge to application skills of language (Wang,2016:985). Thus, the evaluation system should address accuracy, fluency, and appropriateness to generate more positive backwash. In other words, students should not only master language knowledge, but develop their ability to communicate effectively in the real context (Huang, 2019:445).

Assessment Literacy

The concept of assessment literacy (AL) emerged in education during the 1990s. The American Federation of Teachers, National Council on Measurement in Education, and National Education Association (1990) proposed guidelines for teacher competence in educational

assessment. They stressed the need for teachers to be more aware of assessment in all learning environments, not just within classrooms. These proposed guidelines were categorized into two main strands. The instructional use of assessment where teachers should be skilled at selecting, designing, and evaluating assessments that effectively would improve learning, teaching practices, and overall school performance. Then understanding and communicating assessment results as teachers should be able to identify when assessments were used inappropriately and effectively could communicate results to parents, students, and other stakeholders (Mertler & Campbell;2005).

Further research emphasized several additional aspects of AL for educators: understanding assessment quality (e.g. reliability, content validity, fairness), utilizing effective assessment design (close-ended and open-ended tasks), using alternative assessments (portfolios, formative assessments) and applying technology for making more informed decisions about assessment and its role in student learning. AL involved both understanding the core principles of assessment and their actual application in the context of classroom activities as well (Tsagari, 2020; Giraldo, 2018 and Mertler & Campbell,2005).

Consequently, successive research aimed to establish a strong foundation for determining the LA needs of different stakeholders involved (policymakers, teachers-in-training, students themselves, and university instructors. Moreover, Stiggins (1995) questioned the role of teaching and learning in achieving AL. Going beyond simply what and how to assess, the goal should be to develop theoretical frameworks for AL, along with analyzing the specific

needs of each stakeholder group. By doing this, LA can be used to truly improve teaching and learning. However, there is still a gap in answering this question as evidence suggests existing issues in the current pedagogy used to train people in AL (Fulcher, 2020:10).

Language Assessment Literacy (LAL)

Recently, a new area of focus known as (LAL) emerged. LAL built on the general idea of LA but specifically evolved as a unique domain. It investigated LAL of different stakeholders (e.g., teachers, test writers, and professional language testers) focusing on the skills needed to effectively assess language learning. The degree of understanding LAL needed for different stakeholder groups in different test contexts can induce positive consequences from tests (Prastikawati, Lestari and Mokhtar, 2024; Tsagari, 2020; Giraldo,2018, and Tsagari & Cheng,2017).

There is a strong connection between LAL and AL. However, LAL addressed language itself, its complexities, specific skills, alignment between applied language methodologies (e.g. communicative language teaching) and suitable assessment tasks. LAL built on the basics of AL but added a specialized layer of knowledge and skills specifically tailored to effectively assess language learning and learners (Prastikawati, et al.,2024; Tsagari, 2020 and Giraldo;2018)

The ingredients of LAL for language teachers:

Moving beyond the general components of assessment literacy, Inbar-Lourie (2013) proposed a specific set of ingredients that make up LAL as a special toolkit for language teachers, distinct from general assessment literacy despite sharing some common ground. This included:

- a) understanding the social and political roles of assessment and the responsibilities of the language tester and the potential consequences of assessments on students and teachers.
- b) utilizing assessment mechanics to write, administer, and analyze language tests effectively together with ensuring test quality, and reporting results clearly as well.
- c) having proficiency in classroom-based language assessment strategies through using a variety of assessment methods to provide ongoing feedback and measure student progress effectively.
- d) having the ability to match assessment methods with specific language teaching approaches and pedagogies to ensure that assessments accurately reflected what students were learning and how they were learning it.
- e) realizing the differences between formative and summative assessments, internal (teacher-designed) and external (standardized) assessments, and the challenges of ensuring validity and reliability, particularly when assessing authentic language use (pp.27:45).

A Core List of LAL for Language Teachers

A proposed list of LAL for language teachers was based on three central components, introduced by Fulcher (2020), Davies (2008, as cited in Giraldo, 2018) and Inbar-Lourie (2013). The first component was knowledge which served as the essential foundation. This component had three key dimensions. The first encompassed theoretical considerations such as the meaning of validity and reliability particularly in relation to language and language use. The second dimension acknowledged the importance

of familiarity with major issues in applied linguistics (e.g. communicative approaches to language testing) to allow teachers to design assessments that reflect real-world language use. The third dimension emphasized the importance of teachers' knowledge about the specific factors in the teaching context (e.g. student demographics, learning styles, and available resources) to enable teachers tailor assessments for optimal effectiveness.

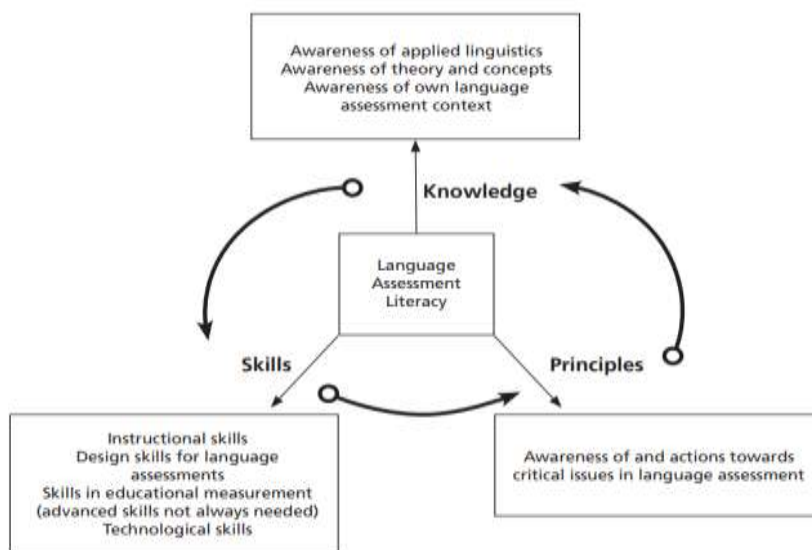
Skills were the second crucial pillar of LAL encompassed five key dimensions. The first was instructional skills which emphasized the effective integration of assessment into daily instruction to guide learning, provide feedback, and promote student progress. The second dimension acknowledged the ability to design effective language assessments for all the four language skills (listening, speaking, reading, and writing) and integrating them seamlessly into assessments. The third dimension focused on the importance of basic measurement skills as quantitative methods for analyzing and interpreting assessment data effectively whereas the fourth-dimension highlighted teachers' meaningful use of these assessment data to inform their teaching practices. The last dimension acknowledged the importance of technological skills to create, administer, and score assessments, which can enhance efficiency and provide valuable data for analysis.

The final component of the LAL framework focused on the paramount principles that guided effective language assessment practices. This component drew on ethics and fairness as essential codes of conduct for all involved in language assessment. It also emphasized the ability to analyze and evaluate existing assessments for their effectiveness and potential biases. Transparency and

student involvement formed other important principles through allowing students to understand the purpose and criteria of assessment together with incorporating their feedback to better enhance their learning experiences. Figure (1) summarizes this core list of LAL.

Figure 1.

A Core List of LAL Dimensions: Knowledge, Skills, and Principles



Source: *Giraldo* (2018: 187)

Test consequences

Testing was designed to serve teaching and learning either in theory or in practice. Testing was not a neutral process which had its consequences for relevant stakeholders whether positive or negative, intended, or unintended which had led teaching to be at the service of testing (Ahmadi, 2016 and Elshaw, et al., 2016). Currently, examinations are used to measure students' abilities and understanding of course content.

Testing and teaching promote and limit each other due to their mutual relationship. Teaching limits testing from macro-objectives to specific ones. Testing also leads to different consequences in teaching concerning teaching objectives, content, methods, and even reliability and validity. Testing and assessment go beyond collecting information about the students' learning and the learning experience. They also invariably affect learners' way of learning, the way the curriculum is delivered, and potentially, the overall learning environment (Huang, 2019 and Muñoz, et al.,2019).

Test consequences was considered a relatively new concept in the domain of language education since the late 1980s. Backwash and impact were two terms commonly used in this field of test outcomes. Impact related to the test effect on individuals in classrooms, schools, or society while backwash related to tests outcomes on both teaching and learning at the classroom level. Thus, many language test makers considered backwash as a dimension of impact (Tzagari & Cheng, 2017 and Ahmadi & Jafari, 2016).

Throughout history, test consequences have been considered from different aspects. At the beginning they were associated with test validity. So, validity evidence at each stage from multiple stakeholder perspectives to better understand the use of test scores in pedagogical and policy practices. In addition to that, an encompassing ethics framework was suggested to examine the consequences of testing on language learning at the classroom, as well as the educational, social, and political levels (Huang, 2019; Muñoz, et al.,2019; Tzagari & Cheng, 2017; Ahmadi & Jafari ,2016 and Wang,2016).

The presence of international students worldwide had led to the need to use language instruction tests, such as English tests, as a standard for language testing. However, it was proposed that skills in each language may suffer impact of such tests on testers and test-takers. In applied linguistics, the effects or consequences of testing were regarded as washback. However, a role reversal has recently occurred in educational settings because of the impact high stake tests exert on different components of teaching and learning process which has altered teaching to be at the service of testing.

Assessment Literacy and washback

Testing determines the way teachers teach and the way students learn as well. This influence can be beneficial or negative. When a test supports effective teaching practices and reinforces the achievement of learning goals, it creates positive washback. Conversely, when a test limits the focus of curriculum and instruction or prioritizes test-taking skills over real understanding or critical thinking, it results in negative washback. Beneficial washback can be achieved if both teachers and students are familiar with the test, its objectives as well as format. This awareness would lead both teachers and students to be ready for tests in a more organized and directed manner.

As a result, previous studies stressed the importance of supporting teachers during the process of test preparation (Elshaw, Heng, Abdullah & Rashid, 2016). This means that for creating positive washback in teaching, some teachers may find difficulty in adapting their teaching strategies to testing processes demands. Hence, these teachers should have assessment literacy to be able to

achieve positive washback effect. In other words, teachers' level of assessment literacy directly impacts their students' learning. Research has shown that teachers dedicate almost 30% of their professional time to prepare assessment tasks for their students, often without a strong background about effective assessment practices. This means that teachers need to be efficient in both what to teach and how to teach it as well. Elshaw, et al., 2016).

In higher education institutions, educators should be familiar with five essential criteria for designing and evaluating several kinds of assessment effectively. These fundamental concepts are according to Brown (2004, as cited in Elshaw, et al.2016:137) are: validity, practicality, reliability, authenticity, and washback effect. This fifth concept assures that assessment literacy is a key factor in having positive effects on learning and teaching (positive washback).

Effective assessment is more than administering tests. Being an assessment literate educator requires possessing specific skills as: setting up clear and measurable learning objectives, utilizing various assessment methods, drawing conclusions about students' work, providing constructive feedback to students, and communicating assessment results to different stakeholders effectively. Hence, this assessment literacy would contribute to the incidences of positive washback of test. Despite the essential role of assessment practices in both the teaching and learning processes, teachers generally postpone assessment issues in later stages of their professional career. In other words, teachers tend to target assessment of learning through measuring students' achievement, rather than assessment for learning which aims for students' learning ability (Elshaw, et al.,2016:139).

Backwash Effect

The beginning of 1990s marked a major focus in the constructs of washback studies for the field of language testing regarding the potential influence of language testing on various aspects of language teaching and learning. In applied linguistics, scholars used various terms to convey the idea of exam effects: “measurement-driven instruction”, “systematic validity”, consequential validity, “curriculum alignment”, “backwash”, and “test impact” (Yi-Ching, 2009:259).

The field of washback effect was advanced by Alderson and Wall in 1993 as a force that necessitates test-takers and tutors to engage tasks or activities due to exams. Then washback studies were directed for years to come empirically investigating the relationship between testing, teaching, and learning determining what test-takers were taught, the way they were taught, what they learnt, and the way they learnt (Alqahtani, 2021 and Dawadi, 2021).

These studies addressed the bidirectional nature of washback as either positive or negative in terms of validity, could have. Positive or beneficial washback (e.g., more motivation, more learning activity) or negative washback (e.g., learning to the test, more learner anxiety, the fear of poor results and the associated guilt, shame, or embarrassment). So, if the focus of language testing is scientific and accurate, it will be able to actively promote the realization of the teaching goal scientifically and effectively or revise it and vice versa is correct. Thus, the main impetus for such studies was to reach a theory to guide testers so that they have the best chance of influencing teaching, education, and society in a positive

way (Alqahtani,2021; Dawadi, 2021; Huang,2019; Suwartono & Riyani ,2019, Ahmadi Safa & Jafari, 2016 and Wang, 2016).

Previous research output shown that washback is a highly complex phenomenon. Research output also assured the interactive multidirectional process of washback due to the varying degrees of complex interaction among the different washback components and participants because multiple stakeholders and multiple factors co-exist within the social context where a test exists. Moreover, research shown that simply changing the content or methods of an examination will not necessarily lead to direct and desirable changes in teaching and learning (Huang,2019; Suwartono & Riyani ,2019, Ahmadi Safa & Jafari, 2016).

Rather various factors within educational contexts are involved leading to desirable washback, e.g., test factors (test methods, test contents, skills tested, purpose(s) of the test), prestige factors (stakes of the test, status of the test) and personal factors (teachers' educational backgrounds and their beliefs). Other factors are micro-context factors (the change or innovation of curricula and teachers' methodologies and the influence of students' learning/school/university setting), and macro-context factors (ranging from government policymaking, school administration, publishing, and general opportunities, to parents' expectations of their children) (Dawadi, 2021; Tsagari & Cheng 2017; Elshaw, et al.,2016; Şenel & Tütüniş,2011and Yi-Ching,2009).

Teachers and washback

Teachers are considered as the prominent factor for success in any educational system. One of the most informative

aspects of the teaching process is assessment procedures as they inform teachers about student learning and help students learn more effectively as well. Therefore, evaluation of students' progress is considered as a key part of teachers' job (Ashraf & Zolfaghari, 2018).

The lacks in teachers' beliefs, knowledge and practice on assessment have negative consequences on student achievement and education. Teachers' inappropriate beliefs and lack of knowledge lead to poor assessment practice. Teachers, who have no explicit views on students' performance, base their teaching methods and objectives on vague standards and lead to inadequate assessment practice. Research has shown that assessment practice lacks several important principles and procedures. Teachers collect a lot of evidence through classroom assessment without using it for instructional decisions. There may also be a problem of translating knowledge and concept into practice by teachers for classroom assessment performance (Najib Muhammad & Bardakçı, 2019).

Considering the possible influence of positive and negative washback of tests at both micro and macro levels, it seems that teachers play an important role in developing different types of washback. In other words, the beliefs and conceptions of the teachers are an important factor in determining the types and intensity of washback effect. Consequently, teachers have become the sources of promoting positive washback (Yi-Ching, 2009).

Innovations to new curricula when introduced may not necessarily lead to a real change of teachers' methodologies and practices unless they have the skills to make appropriate adjustment to new methods. Teachers' role in

promoting the positive side of washback and reducing the negative side is great. As a result, teachers should be provided with adequate training and be familiar with a wide range of teaching methods and assessment techniques as well as have the necessary skills and knowledge to adapt to changes (Muñoz, Véliz-Campos & Véliz, 2019: 105).

Review of Literature and related studies

Theoretical Background

Language assessment skills are interconnected. Studies show strong positive links between all forms of skill-based knowledge, like reading, writing, listening, and speaking (Prastikawati, et al. ,2024:402). This means that as EFL teachers get expertise at evaluating one skill, they naturally improve in assessing others. There's a good reason for this connection. All these skills, despite their differences, contribute to a teacher's overall understanding of language assessment (Language Assessment Literacy). The core principles behind evaluating each skill are quite similar. So, for example, a teacher who becomes an expert at designing reading activities, whether multiple-choice or open-ended, can use that same expertise and build upon it when assessing other language areas.

The backwash effect is the outcome of a test or an examination which results either in a positive or in a negative way. Positive backwash occurs when there is harmony between the teaching and the students' examination or a class test performance. On the other hand, the negative backwash effect happens when teaching and assessment are not aligned. This can occur when the assessment focuses on a narrow range of skills or content, not reflecting what is being taught. These two types of

backwash significantly affect both how teachers teach and how students learn. Furthermore, effective backwash goes beyond just assessment. It also includes open communication between students and teachers about the feedback and evaluations provided. This empowers students to identify their strengths and weaknesses, leading to more effective learning strategies.

Related Studies

Language Assessment Literacy

Assessment literacy was illustrated as considering the knowledge and application of ideal assessment practices. Well-informed teachers in this domain would be capable enough to integrate assessment with teaching. Study in this field went on to distinguish the mismatch between teachers' assessment and evaluation performances and proposed best practice. Studies all over the world revealed that many teachers were inadequately trained to develop, administer, and interpret the outcomes of different types of assessments. Furthermore, the assessment practices employed by teachers did not consistently align with their instructional objectives, and these practices frequently emphasized lower-order cognitive skills. Furthermore, teachers' assessment practices were often not well supported their instructional goals and tended to demand a low level of cognitive processing (Ashraf& Zolfaghari, 2018).

Prastikawati, Lestari and Mokhtar (2024) investigated the level of language assessment knowledge (LAK) among 114 EFL teachers employed in both private and public higher education schools in Indonesia. This study utilized 60 adopted items in a four construct-survey. The findings

revealed that the teachers demonstrated the highest level of expertise in evaluating reading skills, while their proficiency in assessing listening skills was comparatively lower. This means that their overall proficiency in LAK is considerably deficient.

Similarly, Isnawati (2023) conducted a survey assessment literacy of EFL teachers at several secondary schools in an Indonesian context through a set of Assessment Literacy Inventory (ALI). Findings revealed that the EFL teachers' assessment literacy was relatively low, indicating their limited knowledge and skills in assessment.

In another relevant study, Najib Muhammad and Bardakçı (2019) investigated the assessment literacy levels of 101 EFL Iraqi teachers working at secondary and preparatory schools. They used Classroom Assessment Literacy Inventory (CALI) based on the seven standards of teachers' assessment competence for educational assessment of students. The findings revealed that Iraqi teachers got the lowest score from standard seven (recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information). The highest score was from standard four (using assessment results when making educational decisions). Although 77% of teachers reported to have been adequately trained for assessment, the results revealed that they had a low level of assessment literacy.

Ashraf (2017) and Zolfaghari (2018) also investigated the relationship between 120 Iranian EFL teachers' assessment literacy and their reflective teaching. Those teachers were selected randomly from different language institutes to fill out two questionnaires: an assessment literacy inventory, and reflective teaching questionnaire. The findings

suggested the relationship between EFL teachers' assessment literacy and their reflective teaching. Furthermore, it is observed that teachers' assessment literacy predicts their reflective teaching.

Backwash Effect

To explore the factors affecting the washback of the Secondary Education Examination (SEE) English test on students and parents in Nepal, Dawadi (2021) collected data using qualitative approach through 72 oral diaries and 24 semi-structured interviews with students and their parents. The data generated indicated the absence of direct relation between the SEE English test and students' test preparation practices. However, the following several factors were detected to affect the nature of a high-stakes tests' washback; economic implications, social prestige, parents' educational background, family support, and perceived importance of English. She assured that that test washback is linked to the educational and social contexts in which the test is administered.

Muñoz, Véliz-Campos, and Véliz (2019) investigated the washback effect of two different types of assessment procedures on an intact sample of 32 seventh-grade students from a subsidized school in Chile. Through a mix-methods approach, quantitative data were gathered through a self-reported-washback survey administered after both a traditional test and an English language project. Qualitative data, exploring students' perceptions of the test and project's effects, were gathered through a focus group interview. The results showed that both assessment methods were positively valued. However, the focus group interview revealed that the alternative assessment

procedure (the project) was regarded higher in terms of motivation, anxiety, and strategy use.

Saudi Arabia adopts a test-driven educational approach and incorporates the Learning Outcome Based English Language Assessment (LOBELA) in the preparatory year program (PYP) in higher learning institutions. This program is critical as it determines what university programs students would enroll in, directly affecting their future careers. Accordingly, Hazaea and Tayeb (2018) used LOBELA to evaluate the impacts of washback. They discovered that LOBELA, just like other high-stakes exams, critically affects the education settings' teaching mechanisms, teaching staff attitudes and motivation, and the content of assessments.

Huang (2019) attempted to analyze the influence of the English Certificate Examination for Navigation Technology Major on the teaching of English for Navigation Technology Major in China. He highlighted the following aspects to achieve the high reliability of professional English tests and to improve its positive backwash effect: (a) random selection of various sets of questions, (b) improvements of the validity and quality of test questions, (c) adoption of standardized marking, and careful review.

Investigating the washback effect of the high school third grade final exam in the general educational system of Iranian mainstream context was conducted by Ahmadi and Jafari (2016). The study was designed to assess the achievement of high school graduates in different school subjects nationwide exam on EFL teachers' teaching methodology, assessment procedures, and attitudes towards

different aspects of the educational system. The researchers made, validated, and administered a questionnaire to 160 EFL teachers. The results indicated that this exam negatively affected EFL teachers' teaching methodology and increased teaching according to the content and format of the test effect quite noticeably. The results further showed an even stronger adverse effect of the exam on EFL teachers' assessment procedures. However, the teachers' attitude towards different aspects of the educational system was not found to be as strongly affected as the other two variables.

Şenel and Tütüniş (2011) investigated the negative effects of testing and their reasons on 42 students, divided randomly to two groups: experimental and control at the English Preparatory School of Istanbul Aydın University. Firstly, free writing activities were done with both groups. Then, a writing assessment was given to both groups. The experimental group did not know that it was an assessment whereas the control group did. Then, a questionnaire was given to both groups to see their beliefs and attitudes towards writing exams. The results indicate that learners' anxiety level and fear of negative evaluation increase and their performance decreases when they know that it is an assessment. In other words, students with higher levels of foreign language anxiety both expected and received lower grades than their less anxious counterparts.

Commentary

The findings of these studies highlight a critical gap in our understanding of how assessments impact students' learning. While studies on assessment literacy exist, this field remains underdeveloped (Najib Muhammad&

Bardakçı, 2019). Additionally, most washback research focused primarily on high-stakes exams, neglecting the impact of everyday classroom assessments. This presents a valuable opportunity. By improving the language assessment literacy of EFL teachers, particularly regarding positive washback effects, this gap can be bridged. This study has the potential to significantly improve how teachers utilize assessments to enhance student learning, going beyond the pressure of exams.

Context of The Problem

Current research on LAL in EFL field highlight the inadequacy of teacher training programs as well as the absence of sufficient professional development opportunities. Hence, it is important to first explore the EFL teachers' existing language assessment expertise to identify their proficiencies and limitations as a step towards promoting stronger LAL skills (Prastikawati, et al., 2024) and attempting to create a positive backwash effect of their assessment.

As the goal of education in Egypt shifts towards capacity building, teachers will have to help students understand that high academic achievement does not require memorization (Egypt Today, 2018). Instead, teachers will have to develop teaching strategies that support the competency-based approach that the new curriculum defines (UNICEF, 2018). In terms of competency-based assessments, consideration of how to assess students, understand new ideas, solve problems, and retain information will be essential to align the objectives of the curriculum with effective teaching strategies, as well as in

conveying to students that academic achievement will be more broadly defined and assessed.

To validate the problem, the researcher designed and submitted a needs-analysis questionnaire of assessment literacy for EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program in the second semester of the academic year 2021/2022. The objective of the questionnaire was to provide the researcher with insights into gaps in EFL graduate students' assessment literacy in the context of English language teaching to improve their understanding and use of assessment tasks effectively. The questionnaire revealed a significant gap in the students' understanding of assessment, particularly regarding the micro and macro skills of reading and writing. Notably, there was a complete lack of knowledge about assessment practices for listening and speaking skills.

The survey also revealed a clear desire among the EFL graduate students to improve their assessment literacy across all language skills (reading, writing, listening, and speaking). They recognize the importance of designing valid and reliable assessments to provide constructive feedback that enhances student learning. However, while familiar with terms like reliability and validity, none of the students were aware of the concept of "backward design" in assessment.

Many EFL graduate students feel unsure of their ability to assess language skills effectively. This lack of confidence can negatively impact their teaching, as traditional assessment methods often lead to negative washback.

Washback occurs when students prioritize learning how to take tests over actually acquiring language skills. Hence, this study proposes a novel approach: utilizing a skill-based language teaching framework to enhance graduate students' understanding and application of language assessment, fostering a positive backwash effect on their future teaching.

Statement of the Problem

Equipping teachers with a deeper understanding of how assessments influence learning will allow them to design assessments that align with curriculum goals through measuring what is being taught, not just test-taking strategies. Teachers' assessment should go beyond memorization and encourage critical thinking, problem-solving, and higher-order skills. Additionally, clear, and actionable feedback should be presented to enable students to be aware of their strengths and weaknesses and guide their improvement as well. Therefore, this study aimed to improve the language assessment literacy of EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program. This was achieved by implementing a skill-based language teaching approach program, specifically focusing on developing assessments that create a positive washback effect. In other words, the program aimed to equip teachers with the knowledge and skills to design assessments that enhance, rather than hinder, student learning.

Questions of the Study

The present study attempted to answer the following question: How can a skill-based language teaching approach program be designed to develop EFL graduate

students' language assessment literacy and positive backwash effect?

More specifically, the following questions can be branched out:

1. What is the effect of using a skill-based language teaching approach program on developing EFL graduate students' language assessment literacy?
2. What is the effect of a skill-based language teaching approach program on developing EFL graduate students' positive backwash effect?

Aims of the Study

The purpose of this quantitative study was to explore the impact of using skill-based language teaching approach to develop EFL graduate students' language assessment literacy and positive backwash effect.

Hypotheses of the Study

The present study attempted to test the following hypotheses:

1. There would be a statistically significant difference between mean scores obtained by the treatment and the non-treatment groups on the post- performance on the test of language assessment literacy (favoring the treatment group).
2. There would be a statistically significant difference between mean values obtained by the treatment and the non-treatment groups on the post- performance on the scale of backwash effect (favoring the treatment group).

Significance of the Study

The use of skill-based language teaching approach within the field of EFL higher education is hopefully expected to:

- Enhance EFL teachers' deeper understanding of assessing the different language skills effectively.
- Enable EFL teachers to create assessments that go beyond rote memorization and test taking strategies towards measuring actual learning progress.
- Encourage EFL teachers to promote high order thinking skills (e.g. critical thinking, problem-solving and communication skills) to enhance overall language acquisition and create positive backwash effect.
- Support EFL students with meaningful learning through creating a clear link between learning and assessment.
- Provide EFL students with specific and actionable feedback to become aware of their strengths and weaknesses in each language skill.
- Reduce EFL students' test anxiety through becoming more self-confident leading to improved performance and a more positive learning environment.
- Enable professional development providers to offer training and workshops that address language assessment literacy leading to overall professional growth and competence of EFL educators in teaching and evaluation practices as well.
- Provide educational researchers with a rich area of research on backwash effect of testing on teaching and to explore strategies for overcoming its negative backwash to the greatest extent.

Delimitations of the study

The present study was delimited to the following elements:

1. The participants of this study were graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program in the first semester of the academic year

2022/2023.

2. All the participants were fresh in Comprehensive Evaluation Course due to the policy of the professional diploma program in the Faculty of Education, Minia University as this course dropouts or repeaters must rejoin it again in the following year.
3. There is no balance between male and female graduate students at the professional diploma program at the Faculty of Education in both the treatment and the non-treatment groups. This is due to the sex ratio of this population.

Definitions of terms

Skill-based language teaching approach

It is defined procedurally in this study as an approach for teaching assessment that goes beyond simply testing knowledge towards promoting a variety of teaching methods for all language skills incorporating different formats like presentations, group projects, or written essays. It transforms assessment from a passive measurement into an active tool for learning and growth ensuring that students are tested on what they have been taught.

Assessment Literacy

- Webb (2002 as cited in Elshawa, et al. :2016: 137) defines it as the knowledge about how to assess what students know and can do, interpret the results of those assessments, and apply the results to improve student learning and program effectiveness.

Language Assessment Literacy (LAL)

- Giraldo (2018: 180) distinguishes (LAL) from general (LA) by defining LAL as the specific knowledge, skills,

and principles in language testing that relates to language, language use, and language pedagogy.

- Language assessment literacy (LAL) is procedurally defined in this study as a practical knowledge of all aspects of assessment that support language teaching to enable educators effectively tailor instruction to meet the needs and expectations of students, parents, and the entire school community.

Backwash Effect

- According to Pearson (1998, as cited in Yi-Ching, 2009:258), public examinations influence not just student learning, but also the attitudes, behaviors, and motivation of teachers, parents, and learners' themselves. This influence works in a backward direction, shaping teaching and learning decisions throughout the course. That is why it is called washback/ backwash.
- Huang (2019 :553) defines backwash effect as the influence or reaction of testing on teaching (including teaching and learning) in the field of education.
- Alqahtani, (2021:21) defines it as an impact that a test has on the teaching and learning process.
- It is procedurally defined in this study as the unintentional influence that a test has on the way teachers teach and students learn. The test itself is the center point, but its impact goes outwards and backwards, shaping what happens in the classroom.

Method

Research Design

The present study utilized quasi-experimental research design. The pre-post non-treatment group design (Hatch and Farhady, 1982) was used in designing and conducting the study. A treatment group and a non-treatment group

were exposed to pre and post means of getting data. The treatment group was only instructed and trained using a skill-based language teaching approach program while the non-treatment group did not receive such training.

Participants of the Study

Thirty (30) EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program in the academic year 2022-2023 were recruited for this study. The participants were randomly divided into two equal intact groups, treatment, and non-treatment. Homogeneity was established between participants in both groups at the entry level before the intervention as follows:

Age.

All the participants recruited in both groups were aged between 23 and 25 at the beginning of the study.

Linguistic background.

This study involves two groups of participants who share a similar educational background. All participants completed 12 years of English language education, from primary school through secondary school, in Minia Governorate, Egypt. Additionally, they all hold degrees in English education from the Faculty of Education.

Pedagogical background

All the participants in both groups are graduate students at the Faculty of Education and have received formal training in teaching methods.

Instructor

The researcher taught only the treatment group by herself while the non-treatment group was taught by another

instructor. This was done to avoid contamination of the procedures of teaching the non-treatment group and to keep the two groups intact.

Variables of the Study

The independent variable.

The use of a skill-based teaching approach program.

The dependent variables.

The level of language assessment literacy and the level of positive backwash effect.

Pilot Study

The pilot sample of the study consisted of fifteen (15) male and female EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program in the second semester of the academic year 2021/2022. The pilot study lasted for fifteen days and helped in determining the validity and reliability of the study tools.

Instruments of the study

A. Exploratory study instruments

1- Needs Analysis Survey of Assessment Literacy

Purposes of the survey. Determining the current level of knowledge EFL graduate students have in various assessment practices.

Construction of the survey. (a) Reviewing the literature related to the assessment design, implementation, and interpretation, (b) Stating the objectives of the survey, (c) Designing a preliminary survey which consisted of two variables with ten questions, (d) Evaluating the preliminary

form of the questionnaire by a jury of 5 TEFL experts, (e) The final version of the questionnaire after modification consists of six closed-ended questions with pre-defined options.

Validity of the survey. A jury of 5 TEFL experts was asked to approve the validity of the survey. 80% of them approved of its face validity, suitability and appropriacy for the study participants.

Administration of the survey. The survey was administered to EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program to determine their needs and challenges in relation to English language assessment practices.

Results. Analyzing the data received from administering the questionnaire revealed a significant gap in their understanding of effective language assessment. This knowledge gap was particularly clear in the assessment areas of reading and writing. Most of the students' responses showed their struggle to realize all the micro skills (like phonics and grammar) and macro skills (like analyzing text structure) involved in language skills. Even more concerning was the complete lack of knowledge about assessment practices for listening and speaking skills. However, the survey also revealed a positive takeaway. EFL graduate students expressed a strong desire to enhance their assessment literacy across all language skills. They recognized the importance of designing assessments that are not only reliable and valid but also contribute to student learning. The survey also highlighted a specific area for further development: the concept of

"backwash" in assessment. While familiar with terms like reliability and validity, none of the students were aware of this concept. This suggested a valuable opportunity to equip those future educators with powerful tools for considering assessments based on desired learning outcomes and then designing assessments that effectively measure those outcomes. Thus, enabling EFL graduate students to truly benefit their students.

2- Measuring Instruments

1. A Test of EFL Assessment Literacy

Purpose of the test. A test of EFL assessment literacy was designed by the researcher for EFL graduate students enrolled in the methodology section of Minia University's Faculty of Education professional diploma program to assess both their knowledge and application of assessment practices, to ensure equality of the participants in the treatment and non-treatment groups through piloting and to measure the degree of improvement of the participants in both groups on their knowledge and application of the targeted aspects of assessment literacy after finishing the program.

Construction of the test. It consists of two sections (knowledge and application) totaling 60 points. The 40-point knowledge section uses multiple choice to assess understanding of general assessment principles and knowledge of different skill types (performance, micro, macro) across the four English language skills (listening, speaking, reading, writing). The application section (20 points) presents four scenarios where students design assessment tasks for specific skill types (intensive, extensive, responsive, etc.) These responses are graded

against a scoring checklist linked to the program's learning objectives, with one point awarded per multiple-choice question and five per application item.

Instructions for the test. They are written in English. They are brief and easy to understand. They include information about the purpose of the test, the way of recording the answers and the time allowed to complete the test.

Piloting the test. A pilot run was conducted with a mixed-gender group of 15 EFL graduate students enrolled in Minia University's Faculty of Education professional diploma program during the second semester of the 2021-2022 academic year. The pilot test determined an average completion time of 90 minutes, based on the total time taken by all participants divided by the number who participated (15).

Validity of the test.

1. The face validity of the test.

It was determined by submitting it to a panel of five TEFL experts who evaluated it based on three key criteria: clarity of language in the test items, alignment with the program's learning objectives, and appropriateness for the target student population (EFL graduate students). The experts feedback and recommendations were incorporated to refine the test into its final form.

2. The internal consistency of the test.

The same piloting sample (15 EFL graduate students enrolled in Minia University's Faculty of Education professional diploma program) took the test. The internal consistency of the individual items of both section one

(knowledge) and two (application) was calculated as shown in table (1). The values of the correlation coefficient, ranged from (0.788: 0.971), are considered acceptable.

Table (1)
Establishing the internal consistency of the test

The correlation of the skills of each section with the total of that section

Sections	Internal consistency
The total of section one (knowledge)	.788**
The total of section two (application)	.971**

Note. **. Correlation is significant at the level (0.01).

Reliability of the test.

Establishing the reliability of the test was done during piloting. The same piloting sample took the test. The reliability coefficient of the test was determined using Alpha Cronbach coefficient. It ranged from (0.732) to (0.760) for each of the sections of the test and for the total of the sections. The alpha coefficient of the whole test is (0.742) which is considered acceptable as shown in table (2).

Table (2)
The Cronbach Alpha's Reliability Coefficient of the test

The test sections	Means	Variance	Standard Deviation	Alpha
Section 1 (Knowledge)	20.77	10.638	113.178	0.760
Section2 (Application)	7.56	7.685	59.061	0.732
Total	28.33	17.083	291.820	0.742

Note. *. Alpha is significant at the level (0.01).

Cohen et al. (2007:506) point out that the alpha coefficient is considered reliable if they range from 0.70 to 0.90. Thus, the reliability coefficient of the test is considered within the acceptable range.

Item analysis. It evaluates the effectiveness of the items and of the test.

Index of difficulty. Analysis of the responses to individual items was calculated to determine item difficulty for the test. The difficulty of the items is understood as the proportion of the persons who answer a test item correctly. The index of difficulty of this test ranged from 0.36 to 0.78, which is considered acceptable.

Item discrimination. It is the ability of the item to differentiate more knowledgeable students from the less ones. To calculate knowledge, the top scoring students are separated from the bottom scoring students and then their response patterns would be compared. It was found that the items had a positive discriminating power. None of the items had a zero- discriminating power. The power of discrimination of the test ranged from 0.20 to 0.25 which is considered acceptable.

Test time. Time taken by each student was recorded, divided by the whole number of the participants (15) who took the test, which was found to be 90 minutes. Thus, the testing time was 90 minutes.

2. A Scale of Backwash Effect.

Purposes of the scale.

A scale of backwash effect was designed by the researcher for EFL graduate students enrolled in Minia University's

Faculty of Education professional diploma to identify their existing skill levels in different backwash effect strategies and to measure the degree of improvement of the participants in both groups on their abilities in handling language assessment duties after finishing their current skill-based language teaching program. Ultimately, the scale not only gauges the program's effectiveness but also empowers students to pinpoint areas requiring further development in their assessment skills.

Construction of the scale.

The researcher developed a 16-statement backwash effect scale using a three-point Likert format. The scale targets three key backwash effect strategy domains and uses positive and negatively worded statements (statements 2 & 9) to capture a comprehensive evaluation. Participants rate their agreement with each statement on a scale of "basic" (1 point), "developing" (2 points), and "professional" (3 points). This translates to a maximum possible score of 48, detailed in Table 3.

Table (3)
The Domains and Strategies of The Backwash effect scale

No.	Scale Domains	No of items
1	Design strategies	4
2	Content strategies	7
3	Logistics/ Support strategies	5
Total number of items		16
Total score		48

Instructions for the scale.

They are clear and concise, written in English for easy comprehension. They include introductory information about the purpose of the scale, the key backwash strategy domains it assesses, the distribution of the scores on the

response categories ("basic," "developing," "professional"), and the method for recording participant answers.

Piloting the scale.

The researcher piloted the scale with a separate group of 15 EFL graduate students enrolled in the same professional diploma program at Minia University's Faculty of Education during the second semester of the 2021-2022 academic year. These participants were not included in the main study. The pilot test measured the average time to complete the scale, finding it to be approximately 15 minutes. This information helped to determine the appropriate timeframe for administering the scale in the main study.

Validity of the scale.

- 1- The face validity of the scale.** It was determined by submitting it to a panel of 5 TEFL experts to evaluate its clarity, relevance, appropriateness for the target audience of EFL graduate students, wording of the statements, their connection to backwash effect strategies, and their suitability for measuring participants' understanding. Based on the panel's suggestions and recommendations, the scale was revised. Initially, it comprised four domains with 32 statements. This feedback led to a streamlined final version with three domains and 16 statements, ensuring a more concise and focused instrument.
- 2- The internal consistency.** The same piloting sample responded to the scale. The internal consistency of each domain was determined and the correlation between the three domains of the scale and the total scale was determined as shown in table (4). The values of the

correlation coefficient ranged from (.828: .964) and significant at (0.01 level), are considered acceptable.

Table (4)
Establishing the internal consistency of the scale/ The Correlation Between Each domain and The Total Scale

The Domains	Internal consistency
Design Strategies	.964**
Content strategies	.941**
Logistics/ Support Strategies	.828**

*Note. **. Correlation is significant at the 0.01 level (2-tailed).*

Reliability of the scale

Establishing the reliability of the scale was done during piloting. The same piloting sample responded to the scale. The reliability coefficient of the scale was determined using Alpha Cronbach (α) coefficient. It ranged from (0.738) to (0.770) for each of the domains of the scale and for the total of the domains. The alpha coefficient of the whole scale is (.765). It is considered acceptable as shown in table (5).

Table (5)
Establishing the reliability of the scale / The Cronbach Alpha's Reliability coefficient of the scale

Domains	Means	Standard Deviation	Variance	Alpha
Design Strategies	6.51	1.804	3.253	0.749
Content Strategies	11.28	3.505	12.285	0.738
Logistics/ Support Strategies	7.91	2.786	7.761	0.770
Total domains	25.88	7.367	54.269	0.765

*Note. **Correlation is significant at the 0.001 level.*

Scale time. The average time to complete the scale during piloting the scale was found to be approximately 15 minutes. So, the appropriate timeframe for administering the scale was 15 minutes.

The skill-based language teaching approach program (SBLT)

This program proposes a groundbreaking approach to empower future EFL educators with a comprehensive understanding and practical application of language assessment principles through leveraging a skill-based language teaching framework. The shift in focus from generic testing to the building blocks of language proficiency will have a transformative impact on their teaching practices. They would gain confidence in their language assessment literacy and enhance their students' genuine language acquisition leading to a positive backwash effect.

Its aim

The aim of this program is to transform the way future EFL educators approach language assessment to create positive learning environments by cultivating a love for language learning. This transformation takes place on two key levels:

1. Empowering Educators:

- The program bridges the gap in language assessment literacy for EFL graduate students, equipping them with a comprehensive understanding of assessment principles.
- This translates to the practical application of these principles, fostering confidence in their ability to design and utilize effective assessments.

2. Revolutionizing Assessment for Student Benefit:

- By embracing a skill-based approach, educators can create assessments that target specific language sub-skills, promoting genuine language acquisition.
- This shift fosters a positive backwash effect, where

assessment becomes a tool to motivate students and enhance learning.

Its Structure

The SBLT program comprises a five-module structure. The first module dives into the foundational concepts of testing, assessment, measurement, and evaluation, followed by exploring general principles like practicality, reliability, validity, authenticity, and washback. The remaining modules tackle each of the four language skills deeply, breaking them down into micro and macro skills. For each skill, the program explores the five performance categories of each skill (imitative, intensive, responsive, interactive, and extensive), equipping educators with the knowledge and tools to design and score effective assessment tasks using various methods like checklists and rubrics.

The activities

This program involves a mix of individual work, group discussions, presentations, and hands-on application through designing and analyzing assessment tasks. Real-world examples, discussions and activities are employed to address the foundational concepts in the introductory sessions. Concept mapping and group discussions are used for breaking down each chosen skill into its micro and macro skills. Sample tests, quizzes and case studies were utilized to explore the five performance categories for each skill. Then role playing scenarios and discussion questions are applied for designing different assessment tasks for each performance category. Creating rubrics or checklists for practicing scoring techniques are also included.

Its components

The SBLT program utilizes a well-structured, two-hour study session format to ensure a consistent and effective learning experience. Each session follows a standardized template, allowing participants to become familiar with the

flow and maximize their learning. Here's a breakdown of the key components within each session:

1. Foundational Knowledge:

- Glossary: Essential terms and concepts related to the session topic are clearly defined and explained.
- Introduction: The session starts with an overview that sets the context, the content and outlines the key objectives participants will achieve.
- Learning Outcomes: These clearly defined goals specify the specific knowledge and skills participants will gain by the end of the session.

2. The Material:

- Videos: relevant videos to the core theme of the session, enhancing understanding are incorporated into the sessions
- Modal Examples: Practical and relevant examples linked to the session topic are presented to illustrate key concepts and facilitate application.

3. Active Learning and Reflection:

- Discussions: Interactive discussions foster active engagement, encouraging participants to reflect upon and explore the introduced ideas and concepts.

4. Assessing Understanding:

- Formative Evaluation Tasks: Participants get to apply their newfound knowledge through assessments and tasks based on the examples discussed. This allows for feedback and evaluation of their understanding.

5. Consolidation and Extension:

- Module Summary: A concise recap of the session's key points solidifies learning.
- Homework Assignments: Assignments encourage participants to extend their learning beyond the session and delve deeper into the material.

Evaluation of the program

The SBLT program used the following evaluation procedures:

- Throughout each session, teachers used a variety of formative assessments like discussion questions, reflections, case studies, scenarios, and design tasks. These activities allowed for continuous feedback and adjustments to teaching based on student understanding. This ensured students were constantly improving.
- Additionally, summative quizzes were given at the end of each unit. These quizzes measured how well students learned the key concepts covered in that unit. This approach provided a step-by-step evaluation of student progress, ultimately contributing to assessing the entire program's effectiveness.

The construction of the training program has gone through the following steps: reviewing the literature related to the domain of assessment literacy, language assessment strategies that promote a positive backwash effect, stating the general and the specific objectives of all modules and their sessions, preparing the content, submitting the program to 5 TEFL jury members to be evaluated according to the following criteria; statement of items, academic verification of the content, appropriateness of the method and the techniques used for the content and the target audience of the study and applicability of the program.

The Instructional Design of the Study

1. A needs-analysis questionnaire of assessment literacy was developed and administered by the researcher to identify the challenges faced by EFL graduate students in developing assessment skills for their future teaching careers. The questionnaire aimed to gather information

that would help students actively participate in upcoming training and strengthen their language assessment skills.

2. *Pre-testing procedures*

- Pre-testing the participants of both the treatment and non-treatment groups, (N=15), using the test of assessment literacy before the intervention to ensure their homogeneity at the entry level. According to Table (6), the average scores (means) and spread of scores (standard deviations) on the language assessment literacy test were close for both the treatment and non-treatment groups. This is statistically confirmed by a nonsignificant t-value, indicating no difference between the groups at the commonly used confidence levels of 95% (0.05) and 99.9% (0.001). This finding supports the homogeneity of the two groups in terms of their assessment knowledge before the intervention began.

Table (6)

Means, standard deviation, mean difference, t-value, η^2 and effect size on the pre-performance of the treatment and non-treatment groups of the test of language assessment literacy (N=15)

Aspects of comparison	Group	Mean	Std. Deviation	t-value	df	Sig. (2-tailed)
Section1	Pre-treatment.	13.27	3.432	0.514	28	0.611
	Pre - nontreatment.	12.73	2.086			
Section 2	Pre-treatment.	3.27	1.163	0.530	28	0.600
	Pre -nontreatment.	3.07	0.884			
Total of the two sections	Pre-treatment.	16.53	3.182	0.733	28	0.470
	Pre -nontreatment.	15.80	2.210			

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

- Pre-testing the participants of both the treatment and non-treatment groups, (N=15), using the scale of backwash effect before the intervention to ensure their

homogeneity at the entry level. According to Table (7), the average values (means) and spread of values (standard deviations) on the scale of backwash effect were close for both the treatment and non-treatment groups. This is statistically confirmed by a nonsignificant t-value, indicating no difference between the groups at the commonly used confidence levels of 95% (0.05) and 99.9% (0.001). This finding supports the homogeneity of the two groups in terms of the backwash effect before the intervention began.

Table (7)

Means, standard deviation, mean difference, t-value, η^2 and effect size on the pre-performance of the treatment and non-treatment groups of the scale of backwash effect (N=15)

Aspects of comparison	Group	Mean	Std. Deviation	t-value	df	Sig.
Design Strategies Domain 1	Pre- treatment.	5.67	1.589	0.270	28	0.789
	Pre-non-treatment	5.53	1.060			
Content Strategies Domain 2	Pre- treatment.	9.20	1.424	-0.810	28	0.425
	Pre-non-treatment	9.67	1.718			
Logistics/ Support Strategies Domain 3	Pre- treatment.	6.47	0.834	0.609	28	0.548
	Pre-non-treatment	6.27	0.961			
Total of Domains	Pre- treatment.	21.33	2.582	-0.651	28	0.521
	Pre-non-treatment	21.87	1.846			

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

- 3- **The intervention.** Participants at the treatment group were trained using the skill-based language teaching approach program passing through all the previously mentioned steps, activities, and tasks.
3. Post-testing the participants of both the treatment and nontreatment groups, (N=15), using the test of language assessment literacy and the scale of backwash effect after the intervention to compare the results with the

pre-testing results.

- 4- The non-treatment group.** Participants in the non-treatment group received instruction on comprehensive evaluation course using the regular way with no skill-based language teaching approach program intervention.

Findings and Discussion

Verifying study hypotheses

Hypothesis 1

The first hypothesis of the study predicted that there was a statistically significant difference (favoring the treatment group) between mean scores obtained by the participants of the treatment and the non-treatment groups on the post-performance on the test of language assessment literacy. Statistical analysis of the obtained data showed that the treatment group achieved a higher degree of improvement than the non-treatment group on this test as t-value (21.527) is significant at (0.01) level and beyond. Thus, the first hypothesis is confirmed. Table (8) below shows the data obtained to test this hypothesis.

To assess the effectiveness of the skill-based language teaching program in boosting the language assessment literacy of EFL graduate students, statistical analysis utilizing the eta-squared formula (η^2) was employed. Cohen et al. (2007:522) have indicated that an eta-squared value of 0.01 signifies a weak effect, 0.06 represents a medium effect, and 0.14 indicates a large effect. The results, presented in Table (8), revealed a remarkably high eta-squared value of 0.943. This falls well within the category of a large effect, suggesting the SBLT program has a substantial and positive impact on improving

language assessment literacy among EFL graduate students.

Table (8)
Statistical analysis of data obtained by the participants of the treatment and the non-treatment groups on the post- performance on the test of language assessment literacy N=15

Aspects of comparison	Group	Mean	Std. Deviation	t-value	df	Sig. (2-tailed)	η^2	Effect size
Section1	Post-treatment.	35.33	1.877	5.117	28	0.000	0.483	High
	Post-nontreatment.	30.67	2.992					
Section 2	Post-treatment.	22.60	1.298	43.243	28	0.000	0.985	High
	Post-nontreatment	5.20	0.862					
Total of the two Sections	Post-treatment.	57.93	2.314	21.527	28	0.000	0.943	High
	Post-nontreatment.	35.87	3.226					

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

The positive impact of the SBLT program was not limited to just comparing the treatment and non-treatment groups. Analyzing the treatment group's performance itself revealed a significant improvement in their assessment literacy knowledge. Table (9) illustrates a substantial difference between their scores on the pre- and post-assessment of language assessment literacy (t-value = 40.758, significant at 0.01 level and beyond). This statistically significant difference indicates a great improvement within the treatment group. Further strengthening this finding, the eta-squared value (η^2) for this comparison is exceptionally high at 0.983, which again falls under the category of a large effect. This reinforces the conclusion that the SBLT program has a powerful influence on boosting the language assessment literacy of EFL graduate students within the treatment group.

Table (9)
Comparison of pre- and post-test scores within the treatment group
(n=15) for the test of language assessment literacy.

Aspects of comparison	Group	Mean	Mean diff.	Std. Deviation	t-value	df	Sig. (2-tailed)	η^2	Effect size
Section1	Pre-treatment.	13.27	-22.067	3.432	-21.846	14	0.000	0.945	High
	Post-treatment.	35.33		1.877					
Section2	Pre-treatment.	3.27	-19.333	1.163	-42.959	14	0.000	0.985	High
	Post-treatment.	22.60		1.298					
Total of the two Sections	Pre-treatment.	16.53	-41.400	3.182	-40.758	14	0.000	0.983	High
	Post-treatment.	57.93		2.314					

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

Hypothesis 2

The second hypothesis of the study predicted that there was a statistically significant difference (favoring the treatment group) between mean values obtained by the participants of the treatment and the non-treatment groups on the post-performance on the scale of backwash effect. Statistical analysis of the obtained data showed that the treatment group achieved a higher degree of improvement than the non-treatment group on this scale as t-value (14.352) is significant at (0.01) level and beyond. Thus, the first hypothesis is confirmed. Table (10) below shows the data obtained to test this hypothesis.

A statistical analysis was conducted to see if a skill-based language teaching program (SBLT) could improve the positive backwash effect for EFL graduate students. Eta-squared (η^2), a measure of effect size, was used. The results, presented in (Table 10) showed a very high η^2 of 0.880, indicating a large effect. This suggests the SBLT program had a strong and positive impact on boosting the positive backwash effect among EFL graduate students.

Table (10)

Statistical analysis of data obtained by the participants of the treatment and the non-treatment groups on the post- performance on the scale of Backwash Effect N=15

Aspects of comparison	Group	Mean	Std. Deviation	t-value	df	Sig.	η^2	Effect size
Design Strategies	Post – treatment	8.87	0.834	4.867	28	0.000	0.458	High
	Post-non-treatment	6.87	1.356					
Content Strategies	Post – treatment	17.53	0.990	13.652	28	0.000	0.869	High
	Post-non-treatment	10.40	1.765					
Logistics/ Support Strategies	Post – treatment	13.00	1.464	10.311	28	0.000	0.866	High
	Post-non-treatment	7.60	1.404					
Total of Domains	Post-non-treatment	39.40	2.354	14.352	28	0.000	0.880	High
	Post – treatment	24.87	3.137					

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

The positive impact was not just between groups. Analyzing the students' responses who received the SBLT program (treatment group) showed significant improvement within the group itself. Table (11) shows a big difference in their pre- and post-assessment scores (t-value = 20.025, significant at $p < 0.01$). This means their positive backwash effect improved significantly. Further supporting this, the η^2 for this comparison was also very high (0.935), again indicating a large effect. This strengthens the conclusion that the SBLT program has a powerful influence on boosting the positive backwash effect for EFL graduate students.

Table (11)

Comparison of pre- and post-performance values within the treatment group (n=15) for the Backwash Effect Scale.

Aspects of comparison	Group	Mean	Std. Deviation	t-value	df	Sig.	η^2	Effect size
Design Strategies Domain 1	Pre-treatment.	5.67	1.589	-6.908	140.0000	.630		High
	Post-treatment.	8.87	0.834					
Content Strategies Domain 2	Pre-treatment.	9.20	1.424	-18.604	140.0000	.925		High
	Post-treatment.	17.53	0.990					
Logistics/suppose strategies. Domain 3	Pre-treatment.	6.47	0.834	-15.020	140.0000	.890		High
	Post-treatment.	13.00	1.464					
Total of domains	Pre-treatment.	21.33	2.582	-20.025	140.0000	.935		High
	Post-treatment.	39.40	2.354					

Note. **. significant at the 0.01 level (2-tailed)

Note. *. significant at the 0.05 level (2-tailed)

Discussion

The following discussion will compare the present study outcomes with what other literature has found, illuminating consistencies, and disparities, and outlining the broader implications of our study

The study's design forms the basis for interpreting the findings.

The design of this study is particularly distinctive because it considers the way all language skills are connected. This focus on interconnectedness is crucial for understanding how the program ultimately benefited the EFL graduate students. By incorporating this approach, the program improved their knowledge of language assessment (language assessment literacy) to become assessment-

literate educators. It also fostered a positive backwash effect on their own teaching practices promoting effective learning strategies and motivation for future language learners. Hence, the study's distinctive characteristics likely stemmed from a combination of the researcher's pre-defined research questions, the chosen theoretical framework, and the unique nature of the SBLT program.

Post-intervention reflections and observations of participants' reactions provided valuable insights.

After participating in the intervention, participants shared valuable reflections and observations. These centered on several key areas:

- **Deeper understanding of language skills:** Participants reported gaining knowledge about both the micro and macro skills that make up each language skill (e.g., reading comprehension involves skills like identifying main ideas and vocabulary knowledge). This multi-layered understanding likely helped them appreciate the complexity of language acquisition.
- **Assessment versatility:** The program equipped participants with the ability to identify different types of performance categories within each skill (e.g., assessing intensive vs. extensive skills in speaking). This knowledge, together with their newly gained confidence in designing assessment tasks, empowers them to create a wider range of evaluations that target specific learning objectives.
- **Positive Backwash Effect:** The program fostered a positive backwash effect by encouraging participants to critically examine the connections between their teaching methods, classroom practices, and the assessment tasks they choose. This self-reflection allowed them to identify areas where their practices

aligned with their learning objectives and areas where there might be a disconnect. By recognizing these connections and potential misalignments, participants are empowered to adjust their teaching strategies for greater effectiveness, ultimately leading to improved student learning outcomes.

Overall, the intervention appears to have fostered not only knowledge acquisition but also a deeper awareness of how assessment practices can inform and improve teaching strategies.

The interpretation of the findings with reference to the related literature

Assessment practices are teachers' toolbox to allow them to maximize the effectiveness of their instruction. Well-informed teachers in this domain would be capable enough to integrate assessment with teaching. Employing these practices enable teachers to fine-tune the delivery rate of their lessons, evaluate the relevance of course content, guide students' learning, assess the effectiveness of their teaching methods, and prepare their students for success in their overall learning journey. These findings align with previous research by Prastikawati et al. (2024), Isnawati (2023), Najib Muhammad and Bardakçı (2019), Muñoz et al. (2019), and Ashraf and Zolfaghari (2018), who all emphasize the importance of integrating assessment with teaching.

Moreover, the SBLT program tackles another crucial aspect of assessment: ensuring it fosters a positive backwash effect on students' learning. Traditionally, poorly designed assessments can have a negative impact, narrowing student focus and potentially discouraging exploration of the broader range of language skills needed in real-world situations. The SBLT program addresses this

by equipping teachers with the knowledge to design assessment tasks that target a wider range of skills and knowledge. This includes the ability to create tasks that reflect real-world language use (e.g. analyzing authentic materials like news articles, presentations, or debates). Thus, teachers can design assessments that truly measure how well students can understand and use language in these contexts. This shift towards assessments that mirror real-world scenarios ultimately leads to a positive backwash effect, motivating students and pushing them further in their language learning journey. These findings align with previous research by Huang (2019), Muñoz et al. (2019), Hazaea and Tayeb (2018), Jafari (2016), and Şenel & Tütüniş (2011), who all highlight the importance of well-designed assessments that promote positive backwash.

However, the current study's findings regarding positive backwash might not be universally applicable. Dawadi (2021) suggests that simply changing test tasks and content may not always lead to significant shifts in student learning practices. She highlights the difficulty of isolating the test itself from other social and educational factors that can influence learning.

Similarly, Ahmadi (2016) emphasizes the potential positive aspects of tests on language learning and curriculum. He argues that knowing the test format and content can help teachers tailor their instruction and potentially improve student performance. These contrasting viewpoints underscore the need for further research on the multifaceted nature of washback on students and their learning. In addition to exploring the design of assessments, future studies could also examine the role of teacher assessment literacy in maximizing the positive impact of assessments.

Challenges

While the present skill-based language teaching approach program to boost EFL graduate students' language assessment literacy and positive backwash effect offers numerous advantages, applying it presented some challenges.

- 1- Balancing the demands of the program with additional practice assignments and activities was a challenge for graduate students with tightly packed schedules. This was overcome by creating a dedicated virtual platform (what's app). This platform facilitated knowledge exchange through question-and-answer, allowing students to clarify doubts, address misconceptions, and discuss assignments collaboratively, maximizing learning efficiency despite time constraints.
- 2- Graduate students usually come with diverse background knowledge and learning needs. Tailoring activities and assessments to cater to these variations was challenging. This was overcome by the pre-intervention needs-analysis survey to identify existing knowledge levels and their needs to tailor activities and choose appropriate materials ensuring a more effective learning experience for all participants.
- 3- Keeping graduate students motivated and engaged in additional language skill development activities can be tricky. This challenge was overcome by incorporating real-world relevance through case studies and assessment scenarios. Additionally, diverse teaching modes like examples, digital materials, and videos were employed to cater to different learning styles. Finally, reflection activities were included, allowing students to connect with the material and personalize their learning

journey, fostering a more engaging and motivating learning environment.

Limitations

The present study was limited to the following elements:

1. One college (Faculty of Education).
2. One program (professional diploma) at the Faculty of Education, Minia University.

Conclusion

This study revealed that language assessment is not a collection of isolated skills, but rather a unified whole with interconnected parts. This means that all the different types of skill-based assessment knowledge showed strong or positive connections to each other. Thus, as EFL teachers become experts in evaluating one language skill, they are likely to develop expertise in assessing other skills as well. This interconnectedness stems from the shared purpose of all language skills that contributes to overall language assessment literacy (LAL). Moreover, this interconnectedness has a significant implication as it can lead to a positive washback effect. As teachers become more confident and skilled in all aspects of language assessment, it can positively impact their teaching practices and ultimately benefit their students' learning.

Implications and Recommendations

- 1- Educators should consider their students' challenges and interests as well. Then, they should structure their teaching to progress from easier topics to more complex ones. They need to use diverse teaching methods to ensure true understanding, encourage critical thinking and increase their motivation towards learning.

- 2- A well-designed comprehensive evaluation system is the backbone of achieving curriculum goals. It goes beyond simply measuring achievement; it provides valuable feedback that ensures continuous improvement for both teachers and students. Teachers can use this information to make teaching adjustments to improve learning outcomes. Students can use the feedback to develop more effective learning strategies, ultimately boosting their learning efficiency.
- 3- Modifying the structure of traditional, discrete-points final tests and using more classwork and ungraded tests as practice before the actual final tests are conducted, can help in getting positive washback and promoting teachers' language assessment literacy.
- 4- Professional educators who are interested in becoming more assessment literate need to note that merely learning about the subjects they teach and improving their pedagogical skills, are no longer sufficient. They must identify their own assessment practices and know how to use these practices to support and improve their students' learning.

Suggestions for further research

- Investigate how the program impacts specific aspects like interpreting test scores, providing effective feedback, or using assessment data for informed decision-making.
- Explore student perspectives on the program's impact on their learning experience and assessment anxiety.
- Compare the effectiveness of the skill-based program with traditional language teaching approaches in terms of assessment literacy and backwash effect.
- Investigate how the program can be integrated with existing EFL curricula to maximize the alignment between teaching, learning, and assessment.

- Investigate the impact of the skill-based approach on developing specific assessment literacy related to a particular skill, such as writing or speaking assessment.
- Conduct follow-up studies with graduates several years after completing the program to assess the sustainability of their assessment literacy and its integration into their professional practice.
- Adapting the program for different EFL learning contexts (e.g., online learning, specific skill focus).
- This study may be replicated with a larger and more diverse sample of EFL graduate students from different governates in Egypt.

References

- Ahmadi Safa, M., & Jafari, F. (2016). The Washback effects of high school third grade exam on EFL Teachers' Methodology, Evaluation and Attitude. *Applied Research on English Language*, 5(1), 77-98.
- Alqahtani, F. (2021). The Impact of Language Testing Washback in Promoting Teaching and Learning Processes: A Theoretical Review. *English Language Teaching*, 14(7), 21-26.
- Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' Assessment Literacy and Their Reflective Teaching. *International Journal of Instruction*, 11(1), 425-436. <https://doi.org/10.12973/iji.2018.11129a>
- Dawadi, S. (2021). Factors affecting washback of a high-stakes English as a foreign language test. *TESL-EJ: The Electronic Journal for English as a Second Language*, 25(3).
- Egypt Today. (2018). Exclusive interview: Education minister details new system. Available via Egypt Today. <https://www.Egypttoday.com/Article/1/61368/Exclusive-Interview-Education-minister-details-new-system>. Accessed 24 April 2024.
- Elshawwa, N., Heng, C., Abdullah, A., & Rashid, S. (2016). Teachers' assessment literacy and washback effect of assessment. *International Journal of Applied Linguistics and English Literature*, 5(4), 135-141. doi:10.7575/aiac.ijalel
- Fulcher, G. (2020). Operationalizing assessment literacy. *Language assessment literacy: From theory to practice*, 8-28.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional*

- Development*, 20(1), 179-195. <https://doi.org/10.15446/profile.v20n1.62089>.
- Hazaea, A. & Tayeb, Y. (2018). Washback effect of LOBELA on EFL teaching at preparatory program of Saudi University. *International Journal of Humanities and Applied Social Science*, 3(3), 1-10.
 - Huang, X. (2019). The backwash effect of language testing on professional English learning and teaching. *1st International Symposium on Education, Culture and Social Sciences (ECSS 2019)* (pp. 553-558). Atlantis Press.
 - Inbar-Lourie, O. (2013). Language assessment literacy: What are the ingredients. In *4th CBLA SIG Symposium Programme, University of Cyprus*.
 - Isnawati, I. (2023). EFL TEACHERS' ASSESSMENT LITERACY. *LLT Journal: A Journal on Language and Language Teaching*, 26(2), 760-769.
 - Mertler, C., & Campbell, C. (2005). Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory". *Online Submission*.
 - Moustafa, N., Elghamrawy, E., King, K., & Hao, Y. (2022). Education 2.0: A vision for educational transformation in Egypt. In *Education to build back better: What can we learn from education reform for a post-pandemic world* (pp. 51-74). Cham: Springer International Publishing.
 - Muñoz, P., Véliz-Campos, M., & Véliz, L. (2019). Assessment in the English language classroom in Chile: Exploring the washback effect of traditional testing and alternative assessment on seventh grade students. *Paideia Revista de Educación*, (64), 97-118.
 - Najib Muhammad, F., & Bardakçı, M. (2019). Iraqi EFL teachers' assessment literacy: perceptions and practices. *Arab World English Journal (AWEJ) Volume*, 10.
 - Prastikawati, E., Lestari, M., & Mokhtar, M. (2024). Investigating Skills-Based Language Assessment Literacy of EFL Teachers in Indonesia. *KnE Social Sciences*, 394-404.
 - Şenel, E., & Tütüniş, B. (2011). The washback effect of testing on students' learning in EFL writing classes. *Dil dergisi*, (153), 43-52.
 - Suwartono, T., & Riyani, C. (2019). Authentic assessment in ELT: hopes, challenges, and practices. *Refleksi Edukatika: Jurnal Ilmiah Kependidikan*, 9(2).

- Tsagari, D. and Cheng, L. (2017). Washback, Impact, and Consequences Revisited. In E. Shohamy et al. (eds.), *Language Testing and Assessment*, Encyclopedia of Language and Education. Springer International Publishing. DOI 10.1007/978-3-319-02326-7_24-1.
- Tsagari, D. (Ed.). (2020). *Language assessment literacy: From theory to practice*. Cambridge Scholars Publishing.
- Wang, R. (2016). Backwash Effect of Language Test on English Language Teaching. In *2nd International Conference on Arts, Design and Contemporary Education* (pp. 983-985). Atlantis Press.
- Yi-Ching, P. (2009). A review of washback and its pedagogical implications. *VNU Journal of Foreign Studies*, 25(4).

Online References

- Resources for videos of assessment <https://www.britishcouncil.cn/en/exams/aptis/researchers/literacy>

