Gist Education and Learning Research Journal. ISSN 1692-5777.
No. 30 (January - June, 2025). pp. 57-80.

EFL Teachers' Conceptions of Designing Classroom-Based Tests

Concepciones de los docentes de inglés como lengua extranjera sobre la elaboración de pruebas en el aula

Dr. Abebe Tewachew

Debark University, College of Social Sciences,
Department of English Language and Literature, Ethiopia

abebetewachew123@gmail.com

Abstract

An essential component of language instruction is classroom-based assessment, which is used to inform instructional decisions and gauge student progress. The current study explores how EFL teachers visualize developing classroom-based assessments at Debark Secondary Schools in the North Gondar Zone. The study employed a concurrent parallel mixed-methods design to investigate how educators develop their evaluation instruments. The study specifically examined how EFL teachers perceive the creation of teacher-constructed assessments in the classroom. Questionnaires were administered using comprehensive sampling, and samples for document analysis were chosen using Yemane's (1967) technique. The results showed that teachers provided adequate preparation time and communicated test instructions, demonstrating thoughtful engagement in test design. However, the tests frequently lacked thoroughness in their structure because important language testing principles were not consistently implemented, test items were not weighted, and item types were narrowly defined. Given these results, the study suggests that teachers' current methods can be improved through focused instruction in test development, with a focus on aligning their ideas with accepted language testing standards.

Keywords: perceptions, teacher-made tests, test development, content knowledge, pedagogical knowledge

Resumen

La evaluación en el aula es un componente esencial de la enseñanza de lenguas, ya que permite orientar las decisiones pedagógicas y evaluar el progreso de los estudiantes. Este estudio explora cómo los docentes de inglés como lengua extranjera (EFL) conciben el diseño de evaluaciones en el aula en las escuelas secundarias de Debark, en la zona de North Gondar. Se empleó un diseño mixto concurrente y paralelo para analizar los enfoques de los docentes en la elaboración de instrumentos evaluativos. En particular, se examinó la percepción que tienen los profesores sobre la creación de evaluaciones desarrolladas por ellos mismos. Se aplicaron cuestionarios mediante muestreo censal, y las muestras para el análisis documental se seleccionaron utilizando la técnica de Yemane (1967). Los resultados muestran que los docentes brindaron un tiempo adecuado para la preparación y comunicaron con claridad las instrucciones de las pruebas, lo cual evidencia una participación reflexiva en el diseño. Sin embargo, las evaluaciones presentaron deficiencias estructurales: no se aplicaron de forma consistente principios clave de la evaluación lingüística, no se ponderaron los ítems y se emplearon tipos de preguntas limitados. A partir de estos hallazgos, el estudio sugiere que los enfoques actuales pueden fortalecerse mediante una formación específica en elaboración de pruebas, con énfasis en la alineación con estándares reconocidos en evaluación de lenguas.

Palabras clave: percepciones, pruebas elaboradas por docentes, diseño de evaluaciones, conocimiento del contenido, conocimiento pedagógico.

Resumo

A avaliação em sala de aula é um componente essencial do ensino de línguas, pois orienta as decisões pedagógicas e permite avaliar o progresso dos estudantes. Este estudo investiga como professores de inglês como língua estrangeira (EFL) concebem o desenvolvimento de avaliações em sala de aula em escolas secundárias de Debark, na zona de North Gondar. Foi adotado um desenho de pesquisa de métodos mistos, simultâneo e paralelo, para analisar as abordagens dos docentes na elaboração de instrumentos avaliativos. Em particular, examinou-se a percepção dos professores sobre a criação de avaliações elaboradas por eles próprios. Os questionários foram aplicados por meio de amostragem censitária, e as amostras para análise documental foram selecionadas com base na técnica de Yemane (1967). Os resultados mostram que os professores dedicaram um tempo adequado à preparação e comunicaram claramente as instruções das provas, o que evidencia um envolvimento reflexivo no processo de elaboração. No entanto, as avaliações apresentaram deficiências estruturais: os princípios fundamentais da avaliação linguística não foram aplicados de forma consistente, os itens não foram ponderados e os tipos de questões utilizados foram limitados. Com base nesses achados, o estudo sugere que as práticas atuais podem ser aprimoradas por meio de uma formação específica em elaboração de avaliações, com ênfase na conformidade com padrões reconhecidos na área de avaliação linguística.

Palavras-chave: percepções, avaliações elaboradas por professores, desenvolvimento de testes, conhecimento do conteúdo, conhecimento pedagógico.

Background of the Study

Effective language instruction requires the evaluation of language proficiency, and classroom-based testing is a key tool for tracking student development and guiding instructional choices. Classroom examinations, in contrast to standardized assessments, are naturally adaptable, allowing teachers to tailor evaluation procedures to specific curriculum goals and each student's unique needs (Brown, 2004). These assessments provide prompt, context-sensitive feedback, which enables the making of relevant instructional modifications and increases the usefulness of assessments in the classroom (Airasian & Russell, 2008).

The flexibility and capacity to promote significant student engagement through contextually grounded evaluation are what give teacher-constructed tests their educational worth. These tests, which have their roots in constructivist theory, are intended to mirror and strengthen the actual learning process, promoting deeper cognitive engagement and active student participation (Biggs, 1999). Furthermore, formative evaluation and ongoing feedback are crucial for creating an engaging and introspective learning environment, according to modern language assessment theory (Black & Wiliam, 1998). Accordingly, instructors' creation of successful classroom-based assessments is not just a technical undertaking but also a pedagogical necessity that has a direct impact on the caliber of instruction and learning in the context of EFL.

Despite the clear educational advantages of teacher-constructed assessments, creating them can be a challenging task. Numerous educators encounter obstacles, such as time restraints and insufficient professional development, which can erode their confidence and effectiveness in test design (McMillan, 2007). Furthermore, teachers often opt for specific test forms over more comprehensive, learner-centered evaluation techniques due to external demands, particularly those stemming from high-stakes standardized exams (Gordon et al., 2012). Teachers' feeling of efficacy in test development may be significantly impacted by this misalignment between the intended goal of classroom assessment and the realities of instructional environments.

The conversation surrounding teacher-designed assessments continues to revolve around fundamental concepts such as validity and dependability. While reliability refers to the consistency and stability of test results across several administrations and contexts, validity refers to the extent to which a test accurately measures the desired construct (Messick, 1989). Designing excellent tests that produce significant and fair findings requires a firm understanding of these concepts (Harris, 2008). Teachers who have a solid grasp of these ideas are better able to create tools that accurately represent student learning and advance learning objectives.

Feedback systems are equally significant because they play a crucial role in determining the quality of assessments and the learning outcomes of students. When properly included, feedback supports formative learning strategies and fosters a growth-oriented mindset by helping students identify their areas of strength and areas for improvement (Hattie & Timperley, 2007). There are significant differences in the ways that teachers prepare and evaluate tests; some place more emphasis on scoring results, while others concentrate on creating diagnostic instruments that guide instruction and improve student growth.

To enhance teachers' ability to create effective classroom-based assessments, ongoing professional development is crucial. Programs for structured training can provide teachers with the theoretical underpinnings and practical techniques they need to design assessments that meet curriculum requirements and take into account the diverse learning styles of their students (Gordon et al., 2012). Furthermore, teachers' confidence and instructional judgment in the test-development process can be further strengthened by collaborative approaches, including peer review, co-construction of assessment instruments, and the sharing of best practices.

Teachers' ideas about classroom evaluation, however, are not formed in a vacuum; instead, they are significantly influenced by broader cultural and contextual factors. Teachers' perceptions and implementation of test creation are influenced by factors such as institutional mandates, prevailing educational standards, and the socioeconomic realities of students (Wiggins, 2012). Teachers must employ sophisticated assessment techniques that account for students' varying degrees of language proficiency and cognitive development in multicultural contexts, where linguistic diversity and a range of learner experiences are commonplace.

These factors make English language instructors' ideas on creating teacher-made assessments complex and culturally grounded. Examining these ideas helps create more equal and responsive testing procedures while providing insightful information about the lived reality of classroom assessment. Future studies should investigate the interplay between teacher agency, contextual needs, and assessment design in language learning, with a particular focus on how educators address the pedagogical and practical challenges of creating and administering tests.

Statement of the Problem

One of the most important aspects of assessing student progress in English language training is the development and implementation of teacher-generated assessments. Despite their recognized importance, several barriers and diverse beliefs persist among English language teachers regarding the reliability and effectiveness of these classroom-based tools. According to empirical data, many practitioners face challenges such as inadequate professional development in test design, time

constraints, and external pressures from high-stakes standardized tests, all of which erode their self-efficacy and ability to create reliable and psychometrically sound assessments (McMillan, 2007; Gordon et al., 2012).

Furthermore, many teachers struggle to operationalize these concepts in their assessment artifacts even though current theoretical paradigms emphasize the necessity of adapting assessments to learners' proficiency continua and incorporating adaptive assessment strategies into pedagogical praxis. Additionally, research shows that teachers are inclined to prioritize standardized testing procedures above other evaluation approaches that promote student involvement and cognitive development (Black & Wiliam, 1998). This discrepancy highlights a clear contradiction between the theoretical foundations that support the design of complex assessments and the practical limitations faced by teachers in actual classroom settings.

The widespread lack of instructors' understanding of validity and reliability, two fundamental concepts essential for developing sound and practical assessments, represents a significant obstacle. According to empirical research, many teachers lack the necessary training and conceptual understanding of these concepts, which hinders their ability to provide evaluations that fairly and effectively represent their students' learning outcomes (Harris, 2008). The validity and equity of exams are seriously called into question by this epistemic weakness, particularly in schools with a diverse student body. There is a clear vacuum in ongoing professional development frameworks due to the scant attention given to these fundamental test development principles in teacher education programs.

Furthermore, the development of language tests is inherently complex and requires a comprehensive understanding of subject matter expertise, educational philosophy, and the nuances of language testing. As a result, teachers often find it challenging to create assessments of high standards, as their ideas about how to design tests are closely linked to and generally limited by cultural and contextual factors. Bachman and Palmer (2010) bolster this view by claiming that a large number of foreign language teachers demonstrate poor comprehension of language assessment constructs and little cognitive engagement with subject matter information. This shortcoming consequently hinders their capacity to use suitable teaching strategies, resulting in regular departures from accepted testing practices and jeopardizing the validity of evaluations designed by teachers.

Due to misunderstandings and false assumptions about the nature of language teaching and assessment practices, teachers may choose less-than-ideal language evaluation techniques in such situations. Teachers' ability to create valid and reliable assessments in the language education field is often compromised because they frequently fail to comprehend the fundamental principles of effective test construction. Both generally and specifically in the area of language instruction, these issues are very prevalent in Ethiopian educational institutions.

These worries have sparked an increasing amount of scholarly interest in investigating teachers' ideas about language teaching, evaluation, and testing. Numerous studies, including those by Maingi (2015), LaMar (2010), Altaieb (2013), Habeeb (2013), Frewan (2015), and Yidenek (2018), have contributed to this conversation. Maingi (2015), for example, examined how language teachers perceived the use of reading methods in the classroom and found considerable individual variation in theoretical approaches to teaching. Descriptive statistics and independent sample t-tests were also employed in LaMar's (2010) quantitative investigation, which examined middle school teachers' opinions regarding the use of standardized test data at Georgia Southern University. The results showed that administrative leadership provided teachers with time to examine test results and then adjust their lesson plans accordingly.

EFL instructors' opinions of the English language curriculum in Libyan public schools were investigated in Altaieb's (2013) descriptive survey, which used questionnaires and semi-structured and open-ended interviews to gather data. Constrained instructional time for CLT materials, a lack of professional training, insufficient funding, and other contextual issues were the primary causes of the findings, which revealed a discrepancy between teachers' practices and the principles of Communicative Language Teaching (CLT). Similarly, Habeeb (2013) employed questionnaires and informal telephone interviews to investigate teachers' opinions on teaching English as a foreign language in Kuwaiti kindergarten settings. Although most teachers supported exposing young students to English at an early age, their approval of the curriculum's actual content was more ambivalent.

Additionally, Frewan's (2015) study employed semi-structured interviews and classroom observations to gather data on instructors' opinions regarding teaching communication tactics in a few UK universities. The survey revealed that teachers held negative opinions, attributing students' over-reliance on electronics as a barrier to language learning. In support of this, Yidenek (2018) conducted a local investigation that concentrated on the adoption of integrated classroom assessment procedures by EFL teachers in Debre Sina secondary schools. The study, which employed a descriptive survey design comprising questionnaires, interviews, and classroom observations, found that EFL teachers generally failed to incorporate integrated evaluation procedures into their teaching strategies.

Prior research has examined the views of EFL teachers from various angles, including the use of reading strategies, participation in standardized testing, attitudes toward the English language curriculum, teaching communicative strategies, and implementing English as a foreign language. Teachers' opinions about how classroom assessments are created, however, have been mainly ignored in these studies. According to the current researcher, improving students' language proficiency in language instruction requires an awareness of teachers' perspectives on test design. McDonough (2017) supports

this claim by arguing that EFL teachers' perspectives have a significant impact on the creation of successful evaluations, as well as on careful lesson design and classroom procedures.

Essentially, knowledgeable and optimistic teachers create more effective assessments, which have a positive washback effect on students (Wall, 2000; Fox et al., 2007). This emphasizes the idea that following good testing practices requires a thorough understanding of language teaching methodology. Therefore, the focus of this study is to determine how EFL teachers in Debark Secondary Schools in the North Gondar Zone perceive the construction of teacher-made tests. In particular, the investigation aims to answer the following research question: How do EFL teachers feel about creating assessments for the classroom?

Study's Theoretical Underpinnings

Content and Pedagogical Knowledge Concepts of Teachers in Test Development

The professional orientations that educators use to understand and interact with fundamental teaching methods are referred to as concepts. Professionals are expected to acquire contextually aligned understandings that guide their practice in any specialized sector. This suggests that educators need to develop both disciplinary knowledge and pedagogical expertise. An integrated approach that combines methodological soundness with subject-specific competency is necessary to ensure the efficacy of language instruction, as it cannot be achieved by concentrating only on content delivery (Toh et al., 2006). In this sense, the development of reliable, applicable classroom evaluations heavily relies on the knowledgeable perspectives of teachers regarding pedagogy and content.

EFL Teachers' Subject Matter Knowledge (SMK)

The breadth of a teacher's subject matter expertise is a distinguishing feature that raises language instruction to the level of a profession. This requires a thorough understanding of the language itself, allowing teachers to precisely and clearly explain, illustrate, and clarify linguistic topics. According to Roy and Bairagya (2019), this type of knowledge includes the application of metaphors, examples, analogies, and explanatory frameworks that are appropriate for dynamic classroom settings. Both instructional delivery and assessment procedures suffer when teachers lack a solid understanding of the material, resulting in fragmented learning experiences and lower educational outcomes.

According to Faisal (2016), subject matter expertise is a fundamental prerequisite for good teaching, especially when it comes to teaching English. To effectively contextualize and communicate knowledge in a meaningful way, educators are expected to possess a sophisticated understanding of the formal aspects of language, including grammar, phonology, syntax, discourse, writing, and speaking. In this sense, the relationship between student accomplishment and instructor quality highlights the importance of disciplinary knowledge in determining the caliber of test design (Uwatt, 2001). Learner results are statistically significantly influenced by teachers' subject-matter expertise, particularly in assessment. On the other hand, when teachers lack a solid grasp of the material, they often deviate from the fundamental principles of language testing and struggle to create evaluations that are both level-appropriate and cohesive.

This viewpoint is supported by Roy and Bairagya (2019), who contend that thorough content knowledge empowers educators to decide what and how to teach, whether it be in discourse, syntax, phonology, grammar, or the fusion of receptive and productive language abilities. Similarly, Freeman (2002) and Soepriyatna (2012) emphasize the value of topic mastery in helping students overcome conceptual obstacles and promote a deeper understanding of the material.

Additionally, since superficial familiarity frequently leads to problems when creating tests, language teachers must internalize the material they plan to teach. According to Uwatt (2001) and Bisong (2005), subject matter knowledge is an integrated framework that enables educators to find, analyze, and use content in ways that enhance teaching and evaluation. A lack of this knowledge base restricts their ability to provide tests that align with students' competency levels, and it also makes it more difficult for them to respond effectively to student inquiries. Teachers are ill-prepared to develop assessments that offer insightful feedback and encourage student motivation if they lack a solid theoretical understanding of how different language abilities should be assessed.

Therefore, a teacher's capacity to create valid, trustworthy, and pedagogically sound assessments that advance language acquisition is still heavily influenced by their level of subject-matter expertise.

EFL Teachers' Pedagogical Knowledge

Without accompanying pedagogical skill, subject matter expertise alone has little effect on improving the caliber of language training. Effective teaching is based on pedagogical knowledge (PK), which is distinct from, yet equally important as, content knowledge. The fundamental tenets of pedagogical knowledge, as stated by Dörnyei (2009), suggest that mastery of the subject matter must be combined with

the methodological know-how required to transform that knowledge into effective classroom practices. He highlights that pedagogical knowledge encompasses the methods, ideas, and frameworks for instruction that guide educators in planning their lessons (p. 251). Regardless of material proficiency, targeted learning outcomes are not met if excellent pedagogy is not purposefully applied.

The practical approaches teachers use to deliver instruction, from lesson planning and student involvement to classroom management and time allocation, are crucial examples of pedagogical savvy. These operational choices shape the entire learning environment and the efficacy of instruction. Murphy (2008) further explains that pedagogy integrates material and classroom activities into a cohesive and participatory framework, encompassing the dynamic interactions among educators, students, and the educational environment.

On the other hand, the teaching process becomes disjointed and mechanical when pedagogical awareness is lacking or inadequate, even when topic knowledge is good. In these situations, the learning process loses its ability to engage, adapt, and meet the students' needs. Therefore, in EFL contexts, pedagogical knowledge is a crucial factor in determining the quality of instruction rather than an additional asset. Through deliberate, introspective, and student-centered techniques that promote successful language acquisition, educators can empower themselves to realize their subject-matter expertise.

Teachers around the world are increasingly being urged to develop effective classroom management techniques. Such proficiency is necessary to create clear, goal-oriented instructional sequences, maximize instructional time, and maintain a steady pace in the classroom. Having a solid grasp of various teaching approaches is equally crucial and should be consciously incorporated into a teacher's professional toolkit. In support of this, Voss, Kunter, and Baumert (2011) and König et al. (2011) emphasize that teachers proficient in a range of teaching techniques are better equipped to determine when and how to apply their subject-matter expertise in the classroom.

It is crucial to select teaching strategies that align with the subject matter and the desired learning objectives. Structured peer discussions, where students work in pairs to complete activities and then provide feedback to the broader group, may be the most effective way to handle specific course components. On the other hand, some subjects may benefit from demonstration-based learning, especially when explaining abstract ideas or natural phenomena that require tangible examples. When teaching productive and receptive language abilities, including speaking, listening, reading, and writing, as well as their sub-skills, these practices are particularly pertinent since they engage learners' cognitive and psychomotor domains. EFL teachers must use sound pedagogical judgment when selecting the most effective strategy.

Additionally, cooperative techniques, such as the jigsaw approach, can enhance instructional diversity. With this method, students are first divided into groups and given particular subtopics to examine thoroughly. After each group has mastered its subject, its members are redistributed into new groups, each of which is composed of representatives from the initial clusters, as König et al. (2011) explain. Peer instruction and a greater understanding of various subjects are made possible by these new groups' cooperative sharing of expertise. Such approaches demonstrate the usefulness of educational flexibility while also encouraging learner autonomy and interaction.

As a result, EFL teachers who possess a broad range of teaching strategies grounded in both theoretical understanding and real-world experience are better equipped to create language assessments tailored to students' levels of proficiency and context. Due to their methodological flexibility, they can match instructional tactics with test design, ultimately resulting in assessments of student learning that are more reliable and valid.

The integrative nature of pedagogy itself—where pedagogical competence arises from the confluence of a teacher's subject matter expertise with a profound understanding of learners' cognitive needs—justifies the use of appropriate teaching approaches in language training. Delivering content is only one aspect of effective language education; another is modifying instruction to promote internalization and comprehension by learners. Students can interact with and assimilate lesson content more effectively and meaningfully when pedagogical concepts are used with consideration and thoughtfulness.

In this way, pedagogy functions at the intersection of theory and practice. In addition to conceptualizing the construction of knowledge, it entails the deliberate implementation of teaching strategies that suit the cognitive, social, and linguistic settings of learners. Effective pedagogy, according to Persaud (2022), reflects this dynamic interaction, emphasizing that both theoretical underpinnings and real-world applications must inform the teaching process in ways that account for the varied requirements and learning preferences of students.

As a result, teachers have a crucial duty to use teaching methods that successfully connect their subject-matter expertise with the conceptual knowledge of their students. According to König et al. (2011) and Kunter and Baumert (2011), pedagogy includes the methodological strategies teachers use to convert material into approachable, problem-solving-focused learning opportunities. It focuses on how knowledge is acquired and processed by pupils in a meaningful way, in addition to what is taught.

Following pedagogical guidelines is therefore essential for high-quality training and, consequently, for creating successful classroom-based tests. Accordingly, Joint Productive Activity (JPA), Language and Literacy Development (LLD), Making Meaning (MM), Cognitive Thinking (CT), and Instructional Conversation (IC) are the five fundamental characteristics of effective pedagogy listed by Scrivener

(2011). These ideas provide the fundamental frameworks for directing instructional practice and ensuring that instructional methods are learner-centered, reflective, and assessment-aligned.

Methods of Research

The convergent parallel mixed-methods design employed in this study was selected to support the two data collection methods—questionnaires and document analysis—that complemented the study's objectives. This approach was chosen because it enables researchers to prioritize both qualitative and quantitative components equally, perform parallel studies, and then combine the results at the interpretation stage (Creswell & Plano Clark, 2011).

The study was restricted to Grades 9–12 at six secondary schools in Debark, which is situated in the North Gondar Zone. Purposive sampling was employed to select these schools, taking into account the accessibility and specific setting of the study population. EFL instructors from the chosen institutions were the intended participants. A thorough sampling strategy was employed to include all teachers in the survey phase, as they answered closed-ended questionnaire items, given the manageable size of the population (N = 36).

To assess teacher-made assessments, particularly midterm and final exams, document analysis was done in addition to the survey. All grade levels of these exam documents were gathered from each of the six schools. The researcher used Yamane's (1967) formula to calculate the correct sample size for the test documents. To guarantee methodological consistency and dependability in data interpretation, standardized processes were adhered to in the retrieval and classification of the test samples.

$$n = \frac{N}{1+N\,(e^2)}$$

n = the required sample size

e = level of precision

N = total population samples (finite population)

1= unite (a constant value)

Six secondary schools participated in the study, and 1,049 test items from all grade levels (Grades 9–12) were gathered. These items were distributed as follows by grade: 272 in grade 9, 278 in grade 10, 224 in grade 11, and 275 in grade 12. Consequently, the sum of these figures is the entire total number of test items, represented by N:

N=272+278+224+275=1,049 $N = 272 + 278 + 224 + 275 = 1{,}049$

In this regard, $n = \frac{1049}{1+1049\,(e^2)}$

The researcher assumed a 5% value for the level of precision.

$$n = \frac{1049}{1+1049\,(.05)^2}$$

$$n = \frac{1049}{1+1049\,(.05)^2}$$

n= 262

As a result, 262 test items from the teacher-made tests were included in the calculated sample size. The researcher used stratified sampling to distribute this sample evenly throughout the grade levels. Loher (2010) defines stratified sampling as a probability sampling method that involves dividing the population into discrete, non-overlapping, and relatively homogeneous divisions, known as strata. To ensure proportionate representation across the various groups—in this case, the different grade levels—Loher also suggests using a specific method to determine the appropriate sample size within each stratum.

In this manner, $n\,i = \frac{N\,i \times n}{N}$

ni= the required number in each grade level (grade 9-grade 12)

Ni= the total number of question items in each grade level

n= the total required number from all strata

N= the total population of all strata

Thus, let us find the number of questions in each grade level.

a) Number of questions from grade 9,

$$n9 = \frac{N\,9 \times n}{N}$$

$$n9 = \frac{272 \times 262}{1049}$$

n9= 68

b) Number of questions from grade 10,

$$n10 = \frac{278 \times 262}{1049}$$

n10= 69

c) Number of questions from grade 11,

$$n11 = \frac{224 \times 262}{1049}$$

n11= 56

d) Number of questions from grade 12,

$$n12 = \frac{275 \times 262}{1049}$$

n12= 69

All the summation of grade 9, grade 10, grade 11, and grade 12 was obtained in the following way:

n= n9 + n10 + n11 + n12

n= 68 + 69 + 56 + 69

n= 262

The researcher then employed the data saturation method to select question items from every grade level, and the evaluation procedure was repeated until no new themes or issues were identified in the gathered tests. Test items with comparable traits were methodically categorized into matching groups during the analysis. A total of 262 teacher-made exam items were found after this thorough process. The cognitive domains of Bloom's Taxonomy and accepted guidelines for good test design were then used to assess these assessments carefully.

Data Analysis and Interpretation

3.1. Analysis of Questionnaire Data about EFL Teachers' Perceptions

Table 1
The result of descriptive statistics on perceptions of EFL teachers

Items	N	Mean	Std. Deviation
Participating in test development training is a trend that is crucial for constructing efficient tests.	36	3.83	1.13
I find that it is a waste of time to look over the lesson objectives while getting ready for teacher-made tests.	36	2.75	1.46
In test development, it is a good idea for teachers to outline the contents that will be included.	36	3.69	1.28
Teachers' primary responsibility before beginning to develop tests is to take the students' competency levels into consideration.	36	3.91	1.20
Teachers should recap the principles of test before constructing teacher made tests.	36	4.22	.95
Recalling the proper test item selection is important.	36	4.66	.47
It is critical that test content match with learning objectives.	36	3.83	1.32
Using several formats for multiple item tests is a time-consuming process.	36	3.75	1.29
The extensive range of language skills covered in the test indicates a lack of expertise.	36	2.19	1.43
It is essential that teachers review and edit the produced tests.	36	4.16	1.02
Reacting to teacher feedback on the constructed tests is one way to show that you have expertise.	36	4.19	.98

Expected Mean: 3 (5= strongly agree, 4= agree, 3= undecided, 2= disagree, 1= strongly disagree)

The sample mean and the anticipated population mean were compared using descriptive statistics. The majority of items (9 out of 11) scored above the mean, as shown in the table above, providing important insights into teachers' comprehension of test development. According to the statistics, respondents agreed on several crucial

tactics that are essential to creating successful assessments. Notably, the majority of participants emphasized the value of participating in test development training, demonstrating their commitment to enhancing their evaluative skills. Teachers also emphasized the importance of closely aligning test material with learning objectives, demonstrating their commitment to ensuring that assessments accurately gauge the desired results.

On the other hand, items 2 and 9, which were below the mean in Table 1, indicate specific areas that require improvement. Concerns regarding the appropriate coverage of the language skills spectrum and the evaluation of class objectives before test construction are related to these items. These results highlight specific areas where the test development process can be improved. Although there is a generally positive attitude towards test architecture, addressing these flaws could enhance the efficacy and efficiency of test design procedures.

Table 2
Results of One-sample statistics about teachers' perception

Teachers' perception	N	Mean	Std. Deviation	Std. Error Mean
	36	3.75	.41	.07

To determine whether the sample mean obtained from the questionnaire deviated considerably from the expected population mean, one-sample statistics were utilized. The sample mean of the respondents was 3.75, which was higher than the anticipated population mean of 3.00, as shown in Table 2. This shows that the sample mean is greater, but it does not prove that the difference is statistitodita!!!!cally significant on its own.

Table 3
Results of one sample t-test on teachers' perception

perception	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
	10.71	35	.000	.74	.60	.88

Test Value = 3

The one-sample t-test findings are presented in Table 3, which shows a mean difference of 0.74, a significance level (p-value) of 0.000, and a t-value of 10.71 with 35 degrees of freedom. This indicates a statistically significant difference between the sample mean and the population mean, with the p-value being substantially below the traditional threshold for significance. This conclusion suggests that teachers possess a positive and knowledgeable understanding of the stages involved in creating tests, as well as the core principles of effective testing.

Analysis of the Document of Teacher-Made Tests based on Tests Aspects

Table 4
Checklists on the Clarity of Instructions for Teacher-Made Tests

No.	Criteria	Options		Comments
		Yes	No	
1.	Teachers allotted enough time to do the tests.	x		
2.	The instruction typically tells what to do and how to do the items of the test.	x		
3.	The instruction tells the weight of each item.		x	
4.	The test incorporates different item testing formats.		x	

The examined assessments provided students with sufficient time to complete the assigned activities, as shown in Table 4. In particular, the records under evaluation indicated that midterm exams took 15 to 25 minutes and final exams took 40 to 50 minutes, which was sufficient time to complete the test. Item (2) further illustrates that test-takers were given precise instructions that matched the type of questions they were asked. However, as item (3) points out, there was no obvious indication in the testing of the weighting or value given to each test item. Item (4) also shows that the bulk of the prepared evaluations were primarily multiple-choice questions, which limited the range of item types used.

Table 5
Checklists on the Principles of Tests of Stems and Distracters of Teacher-made Tests

No.	Criteria	Options		Comments
		Yes	No	
1.	The stems are authentic.		x	
2.	The test stems incorporate all language skills.		x	
3.	The stems and distracters of the tests are designed in accordance with test fairness.		x	
4.	Each of the stems of teacher-made tests has one correct answer from the objective test formats.	x		
5.	The tests are designed from simple to complex.	x	x	

The developed exams lacked contextual diversity and did not take into account the Ethiopian cultural or national background, as indicated in Table 5, item (1). The majority of the reading passages were taken from other contexts, disregarding their relevance to the pupils' actual surroundings, despite the assessments primarily consisting of objective multiple-choice questions. Regarding integrative testing, item (2) indicates that the midterm and final tests only covered reading, speaking, and grammar and that the assessments did not encompass all language skills.

In terms of test fairness, there was very little evidence of equity in the classroom assessments. As previously mentioned, the fairness of the tests is compromised by their exclusion of full language proficiency and dependence on foreign situations. However, one key feature of successful testing is the presence of items with only one correct response. Furthermore, item (5) indicates that the test construction was not systematic; teachers' irregular sequencing of test problems, sometimes progressing from easy to difficult and at other times in the opposite direction, reflected a poor approach to test design.

Discussion of the Findings

The primary objective of this study was to investigate EFL teachers' opinions on exam design. Teachers generally have a solid grasp of how to create teacher-made assessments, as seen by the questionnaire's overall mean score of 3.75, which exceeded the projected mean of 3.00. According to Davies (2008), teachers who have a favorable attitude toward testing are more likely to collaborate in developing thorough and useful assessments. Although a positive attitude increases the likelihood of creating

high-quality tests, it does not guarantee rigorous adherence to the exacting criteria of language testing (Brown, 2010).

Positive attitudes, by themselves, are therefore not enough; to guarantee the creation of accurate and trustworthy classroom assessments, additional crucial elements, such as educational background, real-world experience, and specialized training, must also be present.

This was corroborated by the document analysis, which showed that the teacher-prepared assessments included clear, well-written instructions and enough time for completion. According to Harris (2008), test takers' performance within the allocated time is greatly influenced by clear instructions. While unclear instructions often cause test takers to become confused and perform worse, well-written instructions allow students to fully comprehend expectations and approach the test with confidence.

Accordingly, the results suggest that instructors' positive attitude toward test creation is reflected in the useful elements of test design, including providing sufficient time and clear instructions, which reduces misunderstandings during exam administration.

However, the investigation also identified serious issues with the test item weighting. Messick (1989) noted that language teachers frequently undervalue the significance of giving test components the proper weights, and this is supported by the fact that many classroom assessments lacked a suitable allocation of value to individual items. Additionally, Messick pointed out that many teachers make the error of assuming that objective examinations alone are sufficient to gauge language proficiency, overlooking the importance of subjective evaluations that are essential for assessing higher-order thinking abilities.

The data from the document analysis revealed that teachers rarely created authentic assessments despite the questionnaire responses indicating a positive attitude toward test preparation. Furthermore, the records demonstrated that not all language capabilities were fully covered in the created examinations, thereby undermining the premise of test fairness. This result is consistent with Wiggins (2012), who noted that even though teachers have a solid grasp of testing principles, they frequently create less authentic tests that fall short of accurately gauging students' actual language proficiency due to issues such as large class sizes, a lack of technology, and inadequate materials.

Consequently, it can be said that good attitudes by themselves are insufficient to guarantee the efficacy of evaluations. According to language pedagogy, assessments should include questions with only one correct response. The prepared tests met this requirement, as shown in Table 5. The test items were arranged erratically and inconsistently, nevertheless. According to language testing experts, such as Brown (2004), test-takers experience anxiety and frustration when presented with challenging questions at the start of an exam. The random order of test items may have had a detrimental effect on students' performance on these assessments.

Conclusions

To triangulate information obtained from document analysis and questionnaire replies, the study used a concurrent parallel mixed design. Percentages means, and standard deviations were used to analyze and present the quantitative data from the surveys in tabular form. In the meantime, checklists based on the principles of language testing were used to collect qualitative data from the document archive.

The combined results from the two datasets highlight the importance of teachers' perspectives on test design in creating successful evaluations. Furthermore, it becomes clear that pedagogical expertise and content understanding are both essential elements that significantly impact test preparation quality. However, other underlying reasons prevent teachers from creating the best assessments; thus, positive attitudes themselves do not ensure the provision of high-quality examinations.

An examination of the documents showed that the tests were not authentic in most cases. The examinations did not outline the weighting of individual test items despite providing ample time for completion and including detailed, unambiguous instructions that reduced student misunderstanding. Additionally, there were only multiple-choice questions on every test. In terms of testing adherence, teachers rarely created contextually appropriate exams. The tests did not fully cover all the necessary language abilities despite being generally well-structured and each question having only one correct answer.

Implication of the Study

The study's conclusions suggest that although EFL teachers exhibit a solid fundamental understanding of test construction principles, focused professional development that prioritizes authentic, skill-integrated, and contextually relevant assessment procedures is still desperately needed. To ensure alignment with legitimate and trustworthy assessment standards, teachers' pedagogical and content knowledge must be strengthened through organized professional development. Furthermore, to overcome real-world obstacles such as large class sizes, scarce resources, and an excessive reliance on specific exam formats, systemic assistance is necessary. By filling in these gaps, classroom-based tests can become more accurate representations of students' actual language proficiency, leading to more equitable and successful outcomes.

References

Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment*. New York: McGraw-Hill.

Altaieb, S. (2013). *Teachers' Perception of the English language Curriculum in Libyan Public Schools: An investigation and assessment of implementation process of English curriculum in Libyan public high schools*. University of Denver.

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford University Press.

Biggs, J. (1999). "What the student does: Teaching for enhanced learning." *Higher Education Research & Development*, 18(1), 57–75.

Bisong, J. (2005). *The Internet and the Language Teacher in Nigeria: A Study of the Impact of Information and Communication Technology on English Language Teaching*: *A Festschrift for A. O. Akinyemi* (pp. 133–143).

Black, P., & Wiliam, D. (1998). "Assessment and classroom learning." *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Longman.

_____(2010). *Language Assessment: Principles and Classroom Practices*. Pearson Education.

Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research. (2nd ed.) Thousand Oaks, CA: Sage.

Davies, A. (2008). *Assessing Academic English: Testing English Proficiency 1950–1989 – the IELTS Solution*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition*. Oxford University Press.

Faisal, A. (2016). *Exploring the impact of testing practices on language proficiency*. *Journal of Language Education and Testing*, *12*(3), 245–262.

Fox, J., Wesche, M., Bayliss, D., Cheng, L., Turner, C. & Doe, C. (2007). *Language Testing, Reconsidered,* Ottawa, University of Ottawa Press.

Freeman, D. (2002). The Hidden Side of the Work: Teacher Knowledge and Learning to Teach. A perspective from North American educational research on teacher education in English language teaching. *Language teaching*, *35*(1), 1–13.

Frewan, A. (2015). *An Exploration of EFL Teachers' Perceptions of the Teaching of Communication Strategies* (Doctoral dissertation, University of Sheffield).

Gordon, D. T., et al. (2012). "The impact of professional development on teachers' assessment practices." *Educational Assessment*, 17(1), 18–32.

Habeeb, K. (2013). *Teachers' Perceptions Toward Implementing English as a Foreign Language at Kindergarten: What Can We Learn from the Case of Kuwaiti Kindergarten Teachers?* University of Arkansas.

Harris, D. N. (2008). "Validating teacher-made tests." *Assessment in Education: Principles, Policy & Practice*, 15(2), 179–197.

Hattie, J., & Timperley, H. (2007). "The power of feedback." *Review of Educational Research*, 77(1), 81–112.

Konigi, H., O'Reilly, S., & Buckner, R. L. (2011). The relationship between syntax and semantics in the processing of complex sentences: A fMRI study. *Journal of Cognitive Neuroscience*, 23(12), 3308–3320. Retrieved: doi:10.1162/jocn_a_00035.

Kunter, M., & Baumert, J. (2011). *The Nature of Teacher Knowledge and Its Impact on Student Learning*. In J. Hattie & E. M. Anderman (Eds.), *International handbook of research on teachers' beliefs* (pp. 247–275). Routledge.

LaMar, T. L. (2010). Middle School Teachers' Perceptions of Using Standardized Testing Data for Instructional Decision-Making. *The educational forum, 74(2), 142–153.*https://digitalcommons.georgiasouthern.edu/etd/370?

Loher, S. L. (2010). Sampling: *Design and Analysis* (2nd ed.) Broos/cole, Bosten.

Maingi, J. (2015). *Language teachers' perceptions and practices in enhancing learners' strategy use for reading proficiency in selected secondary schools in Kenya*. [Unpublished doctoral dissertation at Kenyatta University).

McDonough, S. (2017). *Applied linguistics in language education*. Routledge.

McMillan, J. H. (2007). *Classroom assessment: Principles and practice for effective standards-based instruction*. Boston: Pearson.

Messick, S. (1989). "Validity." In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.

Murphy, D. (2008). *The role of assessment in language education: A practical guide*. Cambridge University Press.

Persaud, A. T. (2022). *The Nature of Science: Teachers' Views and Pedagogies* (Doctoral dissertation, University of Toronto (Canada).

Roy, M. K., & Bairaya, R. K. (2019). *Language Teaching and Testing: Concepts, Methodologies, Tools, and Applications*. IGI Global.

Scrivener, J. (2011). *Learning Teaching: The Essential Guide to English Language Teaching*. Macmillan Education.

Soepriyatna, A. (2012). Investigating and assessing the competence of high school teachers of English in Indonesia. *Malaysian Journal of ELT Research, 8*(2), 38–49.

Toh, G. S., & Tan, J. P.-L. (2006). *The role of test preparation in the context of high-stakes testing: A review of research and implications*. Educational Research Review, 1(2), 99-113.

Uwatt, J. (2001). *Assessing Language Proficiency: Approaches and Methodologies*. *Journal of Language Testing and Evaluation*, *9*(2), 120–135.

Wall, D. (2000). The Impact of High Stakes Testing on Teaching and Learning: *Can this be Predicted or Controlled? Pergamon System, 28,499-509.*

Wiggins, G. (2012). "Assessing student performance: Exploring the purpose and limits of testing." *San Francisco: Jossey-Bass*.

Yamane, T. (1967). *Statistics*: An Introductory Analysis (2nd ed.).New York: Harper and Row.

Yidenek Tadesse (2018). *Exploring EFL teachers' integrated approach to classroom assessment practices: A secondary school in focus*, MA thesis, Debre Birhan University.

Yoon, B. (2009). *The influence of teacher experience on student achievement: An empirical study*. Journal of Teacher Education, 60(4), 368–384.

Author

Dr. Abebe Tewachew is an assistant professor in TEFL at Debark University. He has been providing community services at different elementary and secondary schools. Dr. Abebe's future research interest comprises studies that investigate the complex elements affecting the teaching and learning of second languages, the impact of curriculum design in language education, the impact of policies and structures in language teaching strategies, the roles of L_1 to L_2 learning, critical challenges in language assessment, testing, and evaluation and the psychological impacts of worry on language learners, emphasizing how anxiety hinders engagement and academic advancement in EFL context.

ORCID: https://orcid.org/0009-0000-9594-2807