

Perceptions of EFL teachers to construct teacher made tests

Abebe Tewachew^{1*}, Kassie Shiferie (PhD) ^{2*} and Ebabu Tefera (PhD)^{3*}

^{1*}Bahir Dar University, College of Teacher Education, Department of Language Education

Email: abebetewachew123@gmail.com

Phone number:+251938220175

^{2*} Bahir Dar University, College of Teacher Education, Department of Language Education

Email: kassieshiferie@gmail.com

^{3*} Bahir Dar University, College of Teacher Education, Department of Language Education

Email: teferadugna@gmail.com

Abstracts

Testing is a common practice in language teaching. However, teachers' perception is the most determinant factor to develop effective classroom tests. Thus, this study aimed at examining the perception of EFL teachers to construct teacher made tests in Debark Secondary schools at North Gondar Zone. Concurrent parallel mixed design was employed to investigate the issue. It was ultimately conducted for addressing the research question of 'What is the perception of EFL teachers to construct teacher made tests?' The researcher used comprehensive sampling to take respondents for questionnaire and employed the formula of Yemane (1967) to select samples for the document archive. In this manner, the finding of the study indicated that teachers had good perception to develop tests. Thereby, they provided sufficient time to do tests and clearly put test instructions. However, the weight of each item was not stated and the prepared tests did not comprise different testing items. Additionally, the study found that EFL teachers did not adhere to the principles of language testing. Therefore, the study recommended that teachers need to follow those testing principles along with their good perception.

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

1. Introduction

Testing language skills is a fundamental aspect in language teaching, with classroom tests playing a critical role in evaluating student learning. This test enables teachers to tailor evaluations to the specific needs of their students, aligning closely with curricular goals (Brown, 2004). Classroom tests offer immediate feedback, allowing teachers to adjust their instructional strategies and enhancing the relevance of classroom tests compared to standardized tests (Airasian & Russell, 2008). The significance of teacher made tests lies

in their adaptability and their capacity to foster student engagement. By allowing teachers to design tests that reflect their unique classroom contexts and those tests ensure that evaluations are meaningful and effective. Constructivist approaches highlight the need for tests to mirror the learning process, promoting active student participation (Biggs, 1999). Furthermore, the theories of language tests emphasize the importance of continuous feedback between teachers and students, enhancing the overall learning experience (Black & Wiliam, 1998).

Despite their advantages, constructing effective tests presents numerous challenges. Many teachers encounter constraints such as limited time and insufficient training in developing tests, leading to feelings of inadequacy regarding their testing practices (McMillan, 2007). External pressures, such as high-stakes standardized testing, can also compel teachers to prioritize certain test types over others, impacting their ability to develop comprehensive teacher made tests (Gordon et al., 2012). This disconnect between perceived goals of testing and classroom realities can significantly shape teachers' perceptions about their own efficacy.

Key concepts like validity and reliability are central to the discussion of teacher made tests. Validity pertains to how well a test measures what it intends to measure, while reliability refers to the consistency of test results across different contexts (Messick, 1989). A strong understanding of these principles is crucial for teachers aiming to produce high-quality tests that accurately reflect student learning (Harris, 2008). Teachers who grasp these concepts are more likely to construct effective tests that are fair and equitable for all students. Feedback mechanisms are also vital in the processes of developing tests. Effective feedback helps students identify their strengths and weaknesses, guiding their learning and promoting a growth mindset (Hattie & Timperley, 2007). Teachers' perceptions of test preparation can vary, influencing how they design tests and interpret results. While some teachers may prioritize grades, others focus on providing detailed testing aspects aimed at fostering learners' improvement.

Professional development is essential for enhancing teachers' skills in constructing classroom tests. Targeted training programs can equip teachers with the necessary tools and knowledge to develop tests that meet educational standards while addressing the diverse needs of their students (Gordon et al., 2012). Collaborative practices, such as peer review and sharing best practices, can further support teachers' confidence and capabilities in this area. Cultural and contextual factors also significantly influence teachers' perceptions of teacher made tests. Factors such as cultural attitudes toward test preparation, institutional policies, and the socioeconomic backgrounds of students can shape how teachers approach test construction (Wiggins, 2012). In multicultural classrooms, teachers must consider the varying linguistic proficiencies and experiences of their students, impacting their strategies in constructing effective tests.

To this end, the perceptions of English language teachers regarding the construction of teacher-made tests are complex and influenced by a variety of factors. Exploring these perceptions can provide valuable

insights that inform practices of test development, ultimately enhancing student learning outcomes. Future research should continue to investigate the dynamics of tests in language education, focusing on teachers' experiences as they navigate the challenges of test construction and implementation.

1.1. Statement of the Problem

The construction and implementation of teacher-made tests are pivotal in assessing student learning outcomes in English language teaching. Despite their recognized importance, a range of challenges and varied perceptions among English language teachers persist regarding the effectiveness and quality of these classroom tests. Current research highlights that many teachers experience constraints such as inadequate training in the development of tests, time limitations, and external pressures from standardized testing environments, which can compromise their confidence and efficacy in developing valid and reliable tests (McMillan, 2007; Gordon et al., 2012). Furthermore, while the frameworks of developing tests emphasize the need for assessing learners' proficiency level and adaptability within teaching practices, many teachers struggle to integrate these principles into their tests effectively. Studies suggest that teachers often prioritize standardized tests potentially overlooking the benefits of different approaches that promote student engagement and learning growth (Black & Wiliam, 1998). This gap indicates a disconnect between theoretical frameworks advocating for developing classroom tests and the practical realities faced by teachers in the classroom.

A significant issue lies in the understanding of validity and reliability among teachers, critical concepts that underpin effective test development. Research indicates that many teachers lack sufficient training and understanding of these principles, which affects their ability to design effective tests that accurately measure student learning outcomes (Harris, 2008). This knowledge gap raises concerns about the fairness and equity of tests, particularly in diverse classrooms where students may have varying linguistic and cultural backgrounds. The lack of focus on these fundamental test development concepts in teacher training programs reflects a significant oversight in professional development initiatives.

Furthermore, constructing language test is not a one-fold task, rather it requests the cognition of pedagogy, content knowledge, knowing principles of language test. This is the reason why teachers fail to design qualified tests because the perceptions of teachers are quite influenced by the cultural and contextual aspects to develop teacher made tests. In link to this, Bachman and Palmer (2010) revealed that foreign language teachers lack better cognition to subject matter and perceptions to the language testing aspects. To this effect, they could not adhere to suitable teaching methodologies. This in turns that teachers skip to follow testing principles.

In such circumstances, teachers may focus on unwanted method of language testing because of misunderstanding and wrong perceptions about the nature of language teaching and testing pedagogy. This is to say that teachers miscarry to recognize the principles of test. To this end, they would be unsuccessful

in constructing test in the language teaching discipline. These challenges are common in Ethiopian educational institutions in general and in the discipline of language teaching in particular.

In this vein, there have been demands of canvassing researches towards the perceptions of teachers in language teaching, assessment and testing. Accordingly, there are a number of studies conducted by different researchers (Maingi, 2015; LaMar, 2010; Altaieb, 2013; Habeeb, 2013; Frewan, 2015; Yidenek, 2018). A study by Maingi (2015) delineated those language teachers' perceptions about teaching the use of reading strategies and came up with the finding that there was an individual variations across the teachers' theoretical perceptions about strategies used in the classroom. The work of LaMar (2010) assessed perceptions of middle school teachers regarding the use of standardized testing data in Georgia Southern University. It was quantitative study in accordance of descriptive statistics and independent sample t-tests. In the study, it was found that teachers were provided time by their principals to analyse and plan instruction on the bases of standardized test results.

The perception of EFL teachers to the curriculum of English language in Libyan public schools by Altaieb (2013) was descriptive survey type and the data gathering tools were questionnaire, and open-ended and semi-structured interview. The study showed that there were differences between teachers and the principles of CLT due to teachers' limited time for teaching the CLT materials, teachers' lack of training to CLT, insufficient funding and other factors. Additionally, Habeeb (2013) conducted a study on 'An assessment of the perception of teachers towards implementing English as a foreign language in Kuwaiti kindergarten schools'. He used informal interviews via phone and questionnaire as the tools to collect the required data and the result of the study indicated teachers agreed that children should familiarize English at early stage, but their agreement to the actual curriculum is weaker.

Moreover, the study by Frewan (2015) investigated teachers' perceptions of the teaching of communication strategies in some universities in the United Kingdom. The researcher used semi-structured interview and classroom observation data gathering instruments. In the study, it was found that teachers' perception was not good because students highly depended on electronic devices during their learning. Furthermore, a local study by Yidenek (2018) conducted on 'Exploring EFL teachers' integrated approach to classroom assessment practices at secondary schools of Debre Sina'. It was descriptive survey and he used classroom observation, questionnaires, and interviews as research tools. The result of the study indicated that EFL teachers did not practice integrated assessment in their teaching.

In the aforementioned studies, attempts have been made about EFL teachers' perceptions from different angles: in terms of teaching the use of reading strategies, the use of standardized test, English language curriculum, implementing English as a foreign language and the teaching of communicative strategies. These, the then researchers hardly made endeavors to look perceptions of teachers in terms of test construction. The researcher believes that teachers' perception in test construction is an indispensable

concern to improve students' language proficiency in language teaching. To support this, McDonough (2017) reported that EFL teachers' perception does not only determine the careful planning of lessons, and classroom performance, but also it governs the construction of tests in language teaching. In other words, teachers with good perceptions and practices design better tests and the effect leads beneficial backwash on learners (Wall, 2000; Fox et.al, 2007). This implies that teachers adhere well the principles of testing if they understand better the pedagogy of language teaching.

In this regard, this study focused on examining the perceptions of EFL teachers to construct teacher made tests in Debark Secondary schools of North Gondar Zone. It ultimately addressed the research question of 'What are the perceptions of EFL teachers to construct classroom tests?'.

1.2. Theoretical Foundations of the Study

1.2.1. Teachers' Perception on Content and Pedagogical Knowledge in Test Construction

Perception is the way experts look the world on a particular issues. In any profession, practitioners need to have the compulsory perception that the profession does require accordingly; for example, in the educational arena, teachers should possess pedagogical and content related conceptions. The quality of language teaching is not assured through providing due attention merely teaching the subject matter (Toh et al., 2006).

1.2.2. Content Knowledge of EFL Teachers (SMK)

There are essentialities to say language teaching is a profession. One of these essentialities is content knowledge of the teacher. It is a significant concept in which teachers possess about the content or subject matter. Roy and Bairagya (2019) suggested it as analogies, examples, similes, metaphors and better perception of the topic to be regularly taught in the actual classroom situation. What will happen when teachers lack content knowledge about the subject? If teachers do not have content knowledge, the teaching-learning process is distorted.

In connection to this, Faisal (2016) remarked that teachers' subject matter knowledge is regarded as the prerequisite to teaching where teachers are expected to recognize, perceive the formal aspects of English and subjects such as grammar, phonology, writing, speaking, syntax and discourse. This enables teachers to substantiate content knowledge in context. Subsequently, the relationship between teacher's quality and student's achievement strongly suggests that teachers' subject matter knowledge matters the tests to be constructed (Uwatt, 2001). Teacher's subject matter knowledge exerts a statistically and quantitatively significant impact on student achievement. On the other hand, if teachers do not have better perception about the subject matter knowledge, they do not adhere the required principles of test.

Roy and Bairagya (2019) confirmed that the content knowledge of teachers enlightens what to teach that he/she needs to master which might be syntax, phonology, grammar, written and spoken language use, comprehension as well as discourse. Similarly, Soepriyatna (2012) and Freeman (2002) stated that teachers

must have content knowledge because it would help learners to cope up problems encountered during their learning and understanding the subject matters.

Furthermore, language teachers should internalize each of content to be taught. When teachers do not possess good perception, about subject matter knowledge, they encounter problems of constructing better tests (assessments) because subject matter knowledge is the holistic package that involves getting and utilizing available information that will enhance the teaching and learning of the subject (Bisong, 2005; Uwatt, 2001). Even they face difficulties about questions asked by their learners if they are poor in their content knowledge. They may not theorize how each language skills are tested in accordance with learners' proficiency levels. Therefore, being knowledgeable in the subject area of the course enables teachers to be better in designing assessments, tests, and learners get inputs from the assessment or the tests so that these learners can be motivated to widen their knowledge.

1.2.3. Pedagogical Knowledge of EFL Teachers

Being proficient in subject matter knowledge without delivering the daily lesson through employing effective pedagogy does not bring a significant change on the aspects of language teaching. Pedagogical knowledge (PK) is the most basic element, which should be aggregated in language teaching apart the content knowledge of teachers. According to Dörnyei (2009), the principles of pedagogical knowledge advocate that teachers' subject matter knowledge need to be accompanied by the required pedagogy. Again, he (p. 251) confirmed that pedagogical knowledge is one of the fundamentals that involve the principles or the strategies in which teachers are going to employ during their teaching. The learning objectives of the lesson are not being achieved unless teachers use the appropriate pedagogies.

Language teaching pedagogies are the ways that teachers employ while providing lessons in their actual classroom situation. These include the teachers' methodological applications: the way the lesson to be taught, classroom management, handling students, time managements and other related aspects. Likewise, Murphy (2008) disclosed that pedagogy is the methodological aspect of teachers, which incorporate the interactions between teachers and students and the environment of their teaching-learning and the classroom tasks facilitated in the classroom situations.

Globally, teachers are advised of having the knowledge of classroom management. This enables them for maximizing their allocation of time for each task to teach at a steady pace so that there would be opportunities of creating clear directions of the lesson. It is also perceived that knowledge of teaching methods need to be inculcated in their mind. With a similar vein, Voss Kunter and Baumert (2011) and König et al. (2011) assured that having a command of various teaching methods enable teachers to recognize when and how their subject matter knowledge is put in to practice. Not all teaching (instructional) methods are practicable for all contents of the lesson. Some contents of the lesson may require peer discussion; teachers are providing tasks (activities) on the bases of their sitting beside their peers. Being in

pair, students will complete the task given and each of them are requested to reflect his/her views to the class. Among teaching contents, some of them are applicable for demonstration; the classroom teacher tries to state some natural phenomena. The event of teaching contents maybe practical illustrations so that learners are really exposed to their mental and motor skills; to teach speaking skills, listening skills, reading skills, writing skills and other sub-skills of English, EFL teachers need to employ the appropriate pedagogical knowledge.

The other teaching contents of the lesson are applicable to other methods like, jigsaw method of teaching; the method is carried out that learners are going to be divided in groups and given names in the form of letters. Regard to this, König et al. (2011) stated that teachers provide different topics to each group and inform them to exhaustively discuss and master all the given points; students come up with different views. The new groups are made up of one specialist from each of the original groups. Then, they work together to teach each other the information they learned in their original group. Thus, teachers having various methods of teaching (language pedagogies) are better in concepts of designing appropriate language tests. The reason for employing appropriate teaching methodology is that language teaching pedagogy is an aggregate of the subject matter knowledge of teachers and the learners understating. Students can internalize the daily lesson easily when it is taught through employing different pedagogical principles. Pedagogy is concerned with the fact that both theory and practice should be taken into account in terms of how the knowledge is produced. In this manner, both theory and practice become the key factors in the interplay with pedagogy encompassing the aspects that includes the references of the students in terms of learning (Persaud, 2022). Teachers have duties to use the appropriate instructions in which that the subject matter knowledge of teachers is forwarded towards his or her students. In the same token, Kunter and Baumert (2011) and König et al. (2011) pursued that pedagogy concentrates on how the knowledge of teachers about the teaching content is taught in a meaningful way and how learners easily conceptualize for tackling problems.

Therefore, there are standards to be the pedagogy of language teaching to be effective and this in turns that teachers would be effective in developing better tests. Regarding to this, Scrivener (2011) mentioned five standards of effective pedagogy: Joint Productive Activity (JPA), Language and Literacy Development (LLD), Making Meaning (MM), Cognitive Thinking (CT) and Instructional Conversation (IC).

a) Joint Productive Activity (JPA)

Language teaching is not a lonely activity. The outcome becomes unsatisfactory if a teacher does not collaborate with the concerned bodies, colleagues, even students and administrators. The best method of teaching is the cumulative effect of aggregated agents; it means teaching-learning is not merely the task of students; it is not purely the duties of teachers in the classroom. Teachers, students, course designers, material developers, and administrators should play a significant role to bring effective results in the

educational arena. According to Yoon (2009), teaching language is being effective when novices and experienced teachers are working together and they are motivated and commonly step to goal oriented approaches. This is un-denied that novice teachers can get better content and pedagogical knowledge from those who are experienced in language teaching.

He adds that working collaboratively allows them to conversation, which teaches language, meaning and values shared about their culture and other aspects of teaching language skills between/among the colleagues. Having shared various views with experts lead to getting immediate feedbacks so that they can take remedial actions about their teaching; the teacher adapt the ways how to organizes students in a variety of groupings, such as by friendship, mixed academic ability, language, project, or interests. Teachers have the role of design instructional activities which enable learners to jointly discuss in groups that would have contribution for developing assessments.

b) Developing Language and Literacy across the Curriculum (LLD)

The second standard, according to Scrivener (2011), teachers have various functions; they respond to students' talk, questions, and making an in-flight change. Teachers need to help learners through modeling, eliciting, probing, restating, clarifying, questioning, and praising, as appropriate in purposeful conversation. EFL teachers are expected to connect students' language with literacy and content area knowledge through speaking, listening, reading, and writing activities.

C) Context (Making Meaning) (MM): Connecting School to Students' Lives

It advocates that teaching language should begin at the grass root level on the bases of what learners know. When the teacher tries to teach, provide examples, issues, illustrations need to concentrate on what students know in their home and community; Yoon (2009) noticed that language teaching should not be detached from their culture and their environment. To this end, language teaching does not go beyond their context, social norms; thereby, learners can construct meaning on the bases of their learning environment. The same is true for constructing teacher-made tests.

D) Teaching Complex Thinking (CT)

Teachers need to assist learners for discharging complex thinking during learning language. They should show ways of solving challenges through relating their real life experiences. Dramatic problems with real-life meaning can help students at any level evaluate, revise, and reorganize their conceptual organizations (Kunter & Baumert, 2011). The main role of discharging challenges is not just reaching on the exact answer rather it needs to address learners get in the topic, skills to exhaust their discussion and this leads them to global thinking about the topic learned.

e) Teaching through Conversation (Instructional Conversation)

If teachers deliver the daily lesson through making learners in pair (dialogue), group and panel discussion, students can learn each other. In such circumstance, they would have rooms to discuss with their friends

than the method of teacher-centered. Through dialoging, students can enhance their speaking, listening, reading and writing skills.

Having the concept of PCK, standards of effective pedagogy, teachers are equipped well to construct qualified tests; they can be fully-flagged how effective tests are developed. From this, one can understand that constructing teacher-made English language tests should not be a solitary activity; experienced teachers should share their views, culture, teaching methodologies to the novices in the teaching environment since constructing tests is the reflection of their method of teaching. Therefore, unless teachers perceive well content knowledge, pedagogical knowledge and the standards of effective pedagogy, they could not design qualified tests.

1.2.4. Conceptual Framework of the Study

Conceptual framework is the scheme, which is used to show interrelations of concepts, theories and phenomena in the study. Therefore, the figure below shows the relationship of teaching and testing.

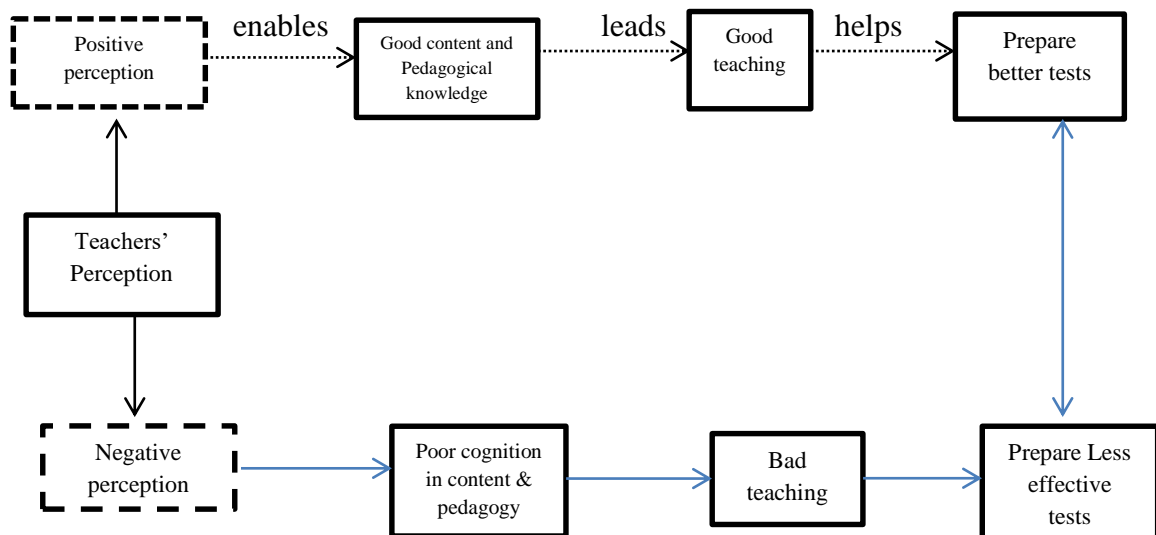


Figure 1: the impact of EFL teachers' perceptions to develop tests, adapted from Brown (2004)

2. Research Methodology

For the sake of conducting this study, concurrent parallel mixed design was employed due to the fact that the nature of the study was accompanied by questionnaire and document analysis. In this context, convergent parallel mixed research is used when researchers are able to balance the methodologies equally, analyze the two components separately, and interpret the two outcomes jointly (Creswell & Pablo-Clark, 2011).

The scope of the study is delimited to six Debarik secondary schools (grade 9 to grade 12) in North Gondar Zone. Purposive sampling was used to select schools in this woreda. The target groups of the study were EFL teachers from these schools. The researcher used comprehensive sampling to take samples of teachers

because they were manageable, i.e., 36 in number. All were filling close ended questionnaire items. The researcher also used document analysis to gather the required data. Thus, teacher made tests of mid and final were collected from six schools. To take samples of tests, the researcher used the formula developed by Yemane (1967). The following procedures were adhered to take the prepared tests from six schools in each grade level.

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

n = the required sample size

e = level of precision

N = total population samples (finite population)

1 = unite (a constant value)

1049 question items were the summation of all grades levels (grade 9-grade 12) from the six secondary schools. These total numbers of questions were collected from each grade as follows:

Grade 9= 272, grade 10= 278, grade 11= 224 and grade 12= 275. Hence, N is the summation of 272+ 278+ 224 + 275, i.e. N= 1049.

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

The researcher supposed 5 % as the value of level of precision.

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

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

$$n = 262$$

Therefore, the required sample size was 262 from the constructed teacher made tests. Owing to this, the researcher found the number of test questions from each of the grade levels. To recognize this, stratified sampling was employed. In reference to this, Loher (2010) stated that stratified sampling is grouped from probability sampling which involves the stratification of population by perishing (dividing) sampling in to none overlapping and relatively heterogeneous group called strata. Apart from this, Loher recommended that researchers can use his formula when they want to know the value of each strata (grade levels) among the groups.

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

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

N_i = 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,

$$n_9 = \frac{N_9 \times n}{N}$$

$$n_9 = \frac{272 \times 262}{1049}$$

$$n_9 = 68$$

b) Number of questions from grade 10,

$$n_{10} = \frac{278 \times 262}{1049}$$

$$n_{10} = 69$$

c) Number of questions from grade 11,

$$n_{11} = \frac{224 \times 262}{1049}$$

$$n_{11} = 56$$

d) Number of questions from grade 12,

$$n_{12} = \frac{275 \times 262}{1049}$$

$$n_{12} = 69$$

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

$$n = n_9 + n_{10} + n_{11} + n_{12}$$

$$n = 68 + 69 + 56 + 69$$

$$n = 262$$

Afterwards, data saturation method was employed for taking these question items from each of grade levels.

The researcher continued referring until getting new issues from the prepared tests. While examining the question items, tests with similar features grouped under the same category. In doing so, the researcher took 262 teacher made tests. The prepared tests were examined in terms of cognitive levels of Bloom's Taxonomy and principles of test.

3. Data Analysis and Interpretation

4.1.2. Analysis of Questionnaire Data about EFL Teachers' Perceptions

Table 1: The result of descriptive statics on perceptions of EFL teachers

Items	N	Mean	Std. Deviation
1. Taking part in test development training is a trend that is crucial to constructing efficient tests.	36	3.83	1.13
2. 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
3. In test development, it is a good idea for teachers to outline the contents that will be included.	36	3.69	1.28
4. Teachers' primary responsibility before beginning to develop tests is to take the students' competency levels into consideration.	36	3.91	1.20
5. Teachers should recap the principles of test before constructing teacher made tests.	36	4.22	.95
6. Recalling the proper test item selection is important.	36	4.66	.47
7. It is critical that test content match with learning objectives.	36	3.83	1.32
8. Using several formats for multiple item tests is a time-consuming process.	36	3.75	1.29
9. The extensive range of language skills covered in the test indicates a lack of expertise.	36	2.19	1.43
10. It is essential that instructors review and edit the produced tests.	36	4.16	1.02
11. 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)

Descriptive statistics is used to compare two means, i.e, the sample mean of the population with the expected mean. As the above table depicted, the majority of the items (9 out of 11) were above the mean. It becomes clear by examining teachers' perception on test development that a number of important themes

come to light. Teachers who were polled appear to agree on a few key techniques that are essential for constructing good tests. Interestingly, most participants concur on how important it is to take part in test development training. Teachers also stress how important it is to match test material to learning objectives, demonstrating their dedication to making sure tests fairly represent learning objectives.

Conversely, items 2 and 9 in Table 1, which received below-average scores reveal about going over the lesson objectives before constructing tests, as well as worries about the range of language skills, which were addressed. This point up places where test development should be strengthened. Despite a general shift toward positive perception, resolving these issues could improve the efficacy and efficiency of test development methods overall.

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

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

One sample t-test is used to determine whether the sample mean of the questionnaire is different from the expected mean of population. Accordingly, in Table 2, the respondents' sample mean was 3.75, but the expected mean of the population was 3.00. This implies the sample mean was greater than the population mean. However, this does not tell the significant differences of the means.

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

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

Test Value = 3

Table 3 shows the sample mean analysis of a one sample t-test ($t = 10.71$, $(df) = 35$, $(sig.) = .000$ and mean difference = .74. This indicates that the p value is less than the anticipated sig. value. Therefore, there was

statistical significance between the sample mean and the population mean. This implies that teachers had good perception to the phases of test construction and the principles of test.

3.1. 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	
		Yes	No
1.	Is there is sufficient time given from each of the tests?	x	
2.	Does the test items' instruction clearly instructs test takers what should be done to each of the question prepared?	x	
3.	Does the weight of each testing items clearly set from the prepared tests?		x
4.	Do the prepared tests encompass various testing items?		x

According to the table above, in Table 4, the prepared tests contained sufficient time , which is enough to accomplish the test items. As the referred document showed, students were allowed to use 15-25 minutes tests for mid-term tests to do the tests. As well, 40-50 minutes were allowed to do their final exams. Item (2) from the above table instructs well test takers according to the nature of questions. Nevertheless, on base of item (3) those tests did not encompass the weight that how many of values contain each test items. Apart from this, majority of the prepared tests were merely accompanied multiple choice items as was shown in item (4).

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

No.	Criteria	Options	
		Yes	No
1.	Are the stems of the tests contextualized?		x
2.	Do the stems of tests involve the different language skills?		x
3.	Are the stems and distracters of the tests developed in aligned with test fairness?		x
4.	Do the constructed classroom tests contain a single right answer?	x	
5.	Are classroom tests organized from ease to difficult /difficult to ease?	x	x

As noted in Table 5 from item (1), the prepared tests did not accompanys different contexts. They even did not comprise the country context. Even though the tests contained multiple choice items (are objective in nature), most of the passages were adopted from the context of other country. They did not demonstrate Ethiopian context. In the context of integrative testing, according to item(2), the tests did not involve all language skills. Reading, speaking and grammar were incorporated from mid and final exams.

With regard to test fairness, classroom test seldom to comprise test fairness. This is due to those did not incorporate all language skills and even as stated in item (1), the tests were directly adopted from the context of other country. Thus, tests did not fulfill tests fairness. However, they contained a single right answer. This is one of a manifestations of effective tests. From item number (5), teachers were organizing tests haphazardly. This is mean to say that they arrange test items sometimes from ease to difficult and sometimes from difficult to simple.

3.2. Discussion of the Results

The purpose of this study was to address the question of ‘What is the perception of EFL teachers to construct tests?’ Therefore, the grand mean (3.75) of the questionnaire items about their perception was greater than the anticipated mean (3.00). This pointed out that teachers had better awareness to construct teacher made tests. According to Davies (2008), teachers with positive perceptions are more likely to work together to develop well-rounded tests. Although it greatly raises the likelihood, a positive insight toward test preparation, it does not ensure adherence to the standards of language testing (Brown, 2010). Consequently, positive perceptions alone are not enough to develop valid tests. Additional requirements include educational background, experiences and other relevant determining factors are the most important elements, which should be thought in order to develop classroom tests.

With this respect, the data obtained from document archive demonstrated that the prepared tests comprised sufficient time to accomplish the tests and their instructions were stated in a well manner. In this connection, Harris (2008) delineated that examination instructions are the determinant factors in which tests takers would like to gain or lose in accordance with the allotted time. If instructions are well stated in the tests, students would have the opportunity to understand very well and can do better; tests test takers will be confused when test instructions are vague. Therefore, we can deduce that since teachers had good perceptions to construct tests, the prepared tests contained better instructions and enough time was allotted so that students will not be able to be confused by instructions and they can do the tests within a given time. In regard to the weight of tests and testing items, classroom tests lacked these essentialities. This finding is congruent with the finding of Messick (1989) that language teachers often lose to put the value of each test and even they have misconception about preparing tests. He added that most teachers believe that to measure students’ language proficiency, objective tests are the most effective tools. However, subjective tests are used to assess learners’ higher order thinking levels.

Even though from questionnaire items, teachers were found that they had positive perception, the data from document archive revealed that they seldom develop authentic tests. Furthermore, the document indicated that the prepared tests were not involving all language skills. This does not assure tests fairness. The result is found from the study of Wiggins (2012). Despite teachers good cognition to testing principles, they hardly prepared less authentic tests, which do not assess test takers’ real language proficiency due to large

class size, lack of technology and materials. Thus, one can deduce that positive perceptions of teachers alone do not confirm the effectiveness of tests. In language pedagogy, it is recommended that tests need to contain one correct answer. As indicated from Table 5, the prepared tests had a single right answer. However, the arrangement of those tests was random. Experts, in language testing, like Brown (2004) claimed that test takers are irritated and anxious when they face by challenging questions from the very begging of test pages. Therefore, students were doing well in those tests since teachers organized them haphazardly.

4. Conclusions

The study used concurrent parallel mixed design for substantiating the data obtained from questionnaire items and the document archive. From the questionnaire, the researcher collected quantitative data which were described in percentage, mean and standardization through using tables. However, the document archive was used to gather qualitative data via using checklists of language testing principles.

The obtained result from questionnaire items and the document archive integratively confirmed that teachers' perception is the most determinant factor to develop effective tests. Their content and pedagogical knowledge also are the most crucial aspect, which play prominent roles in test preparation. This however, the positive perception of teachers alone does not assure the quality of effective tests. There are supplementary issues that lead teachers not to develop better tests.

The document revealed that the constructed tests lacked authenticity. It was found that the prepared tests contained sufficient time and clear and precise instructions so that students were not confused to understand those instructions, but there were no mentioned the weight of each test and all the tests were accompanied by a single testing item, i.e., multiple choice item. With regard to principles of test, teachers rarely developed contextualized tests. Along this line, the tests did not incorporate all language skills regardless of their proper arrangement though there was a single right answer for each question.

5. Implications

Teachers should perceive well subject matters and the pedagogical aspects of language teaching. And they are expected to make better alignments between their content and pedagogical knowledge and the language testing principles. These alignments can assure the effectiveness of tests so that students' language skills can be properly assessed.

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. and 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 the use of 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.) Brooks/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 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' an 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.