



## The Use of Assessment for Learning Rather than Assessment of Learning in EFL Context

Mohammad Zohrabi <sup>1\*</sup> , Fereshteh Nasirfam <sup>2</sup> 

<sup>1</sup> Associate Professor, Department of English Language, Faculty of Persian Literature and Foreign Languages, University of Tabriz, Tabriz, Iran

<sup>2</sup> M.A. in TEFL, Department of English Language, Faculty of Persian Literature and Foreign Languages, University of Tabriz, Tabriz, Iran

Received: 2023/12/30

Accepted: 2024/02/04

**Abstract:** Most EFL schools, both government-run and private, use materials developed based on the curriculum developed by the Ministry of Education. The Ministry of Education in Iran started to update English textbooks for junior and senior high schools to enhance students' communicative competence. To achieve this, there is a need for changes both in methodology and content of the textbooks in assessment methods. This study aimed to investigate whether teachers' tests align with the changes in textbooks or assess students' knowledge of grammar and vocabulary, regardless of their communicative competence. A mixed-methods design was employed, combining quantitative data collected through questionnaires, teacher-made tests, textbooks, and qualitative interview data to provide a comprehensive understanding of the match. Thirty Iranian English teachers from different provinces participated in the study and 100 teacher-made tests were compared with the content and purpose of the updated textbooks. The results analyzed through chi-square test and thematic analysis showed no match between the theory of the tests and the textbooks and revealed that teachers' attitudes towards change of assessment tools contribute to the development of communicative skills.

**Keywords:** Traditional Assessment, Alternative Assessment, Linguistic Competence, Communicative Competence.

---

\* Corresponding Author.

Authors' Email Address:

<sup>1</sup> Mohammad Zohrabi (mohammadzohrabi@gmail.com), <sup>2</sup> Fereshteh Nasirfam (nasirfam.fereshteh@yahoo.com)



## Introduction

Assessment is a very crucial factor in educational settings, particularly in the EFL context. It plays an important role in the process of both teaching and learning and determines the extent to which the goals of education have been achieved. It generally consists of two different types: traditional and alternative assessment. Traditional assessment merely seeks to evaluate whatever the students learn throughout the course of study (Nasab, 2015). However, due to the fact that in this type of assessment, students are passive and the focus is on the instructor and grammatical properties of the language, it is not a proper and effective way to evaluate knowledge. On the contrary, alternative assessment refers to real-world and authentic methods and techniques in the educational field that might be incorporated into regular classroom activities (Hamayan, 1995). In other words, alternative assessment refers to ongoing evaluation procedures that take place in or out of the classroom context at various times. Struyven et al. (2005) assert that there is a strong connection between students' approaches to learning and how they tackle homework and exams during class hours. Due to the fact that assessment has a significant impact on how learners approach learning, a paradigm shift has occurred from "testing learning of students to assessing for students learning" (Birenbaum & Feldman, 1998, p. 92).

To ascertain that the learners are making momentous and perpetual progress in their learning process, teachers should design activities that engage each and every one of them (Gilani et al., 2021). For this purpose, they need to provide learners with different tasks and activities that engage them in producing meaningful utterances instead of just mastering the linguistic properties of the language (Zohrabi & Bimesl, 2022). Therefore, learners need to develop their communicative competence besides their linguistic competence so as to be both fluent and accurate speakers. This can be fulfilled through the employment of different types of alternative assessment methods, which focus on testing the learning capabilities of the learners. Therefore, teachers should design effective tests that assess students' English proficiency in a communicative way and not just focus on grammar and vocabulary.

According to the recent changes made by Iran's Ministry of Education in English textbooks for junior and senior high school, the focus has shifted from learning grammar and vocabulary to improving learners' all four language skills. However, regardless of these changes, students still have difficulty developing their communicative skills and cannot transfer their knowledge to real-life situations. This is because teachers are still following the traditional way of teaching and testing with the focus of developing learners' linguistic competence without any effort to improve their communicative competence. In other words,

the content of textbooks and methodology are not adjusted in a way that they align with the defined educational purpose. Since teaching and testing are inseparable and teachers use tests mostly to check their learner's progress or performance, every change in methodology requires changes in assessment methods as well. With this in mind, the present study investigates the extent to which the tests used by the teachers during or at the end of the semester, include all the skills and sub-skills that are emphasized in the updated version of textbooks and adhere to communicative competence testing methods.

## **Review of the Literature**

### ***Assessment in Education***

According to Dhindsa et al. (2007, p. 1261), assessment is "a systematic process for gathering data about student achievement" and is regarded as a crucial aspect of instruction. An important point to bear in mind is that testing and assessment should not be considered the same. Assessment is the process of collecting information informally on students' current state of knowledge through employment of a variety of methods at different times and in diverse circumstances and contexts (Baker, & Riches, 2018; Berry et al., 2019). On the other hand, testing is a formal and standardized process of evaluation, which provides results based on the activities students have completed. It is mostly implemented in a set time and on a single occasion and is the only appropriate way to gauge how well students are learning (Giraldo, 2020, 2021; Harding & Brunfaut, 2020). Nowadays, the fact that there is only one way to obtain information about students' learning is rejected by many scholars (e.g. Braun et al., 2019; Butler et al., 2021; Fulcher, 2020; Isbell et al., 2023; Jiang et al., 2022; Kim et al., 2020; Kremmel & Harding, 2020; Lee & Butler, 2020; Lee et al., 2021; Levi, & Inbar-Lourie, 2020; Ockey et al., 2023). As Kulieke et al. (1990) declare, testing is viewed as only one component of the vast concept of assessment. Since assessment has a significant impact on how learners approach learning, there has been a shift from "testing learning of students to assessing for students learning" (Birenbaum & Feldman, 1998, p. 92). Each will be briefly explained below.

### ***Assessment of Learning***

Being founded on behaviorist theory, this type of evaluation is most common in educational settings. Behaviorists such as Skinner, Watson, Thorndike, and Pavlov, believe that "learning is a change in observable behavior caused by external stimuli in the environment" (Skinner,

1974). Assessments of learning are usually summative and take place at the end of a unit or course of study in the form of questions derived from the content covered in the class. They are usually meant to attest to learning and inform students about their academic achievement. Typically, this is done by indicating a student's relative position in relation to other students. In this case, the results are typically reported to parents in the form of grades (Marzano, 2006).

### *Assessment for Learning*

Assessment for learning “entails an ongoing process of gauging and monitoring learners’ learning to find the pitfalls and to take wise steps in order to enhance their learning” (Nourdad, 2022, p. 69). The theoretical basis for such an assessment is constructivism, which examines how students construct their own knowledge. The knowledge is constructed based on their interaction with the world and personal experiences. Stavredes (2011) asserts that to internalize new knowledge and information, the learner gives it meaning based on previously-shaped attitudes, beliefs, and experiences. In this situation, the teacher plays the role of a facilitator, who actively cooperates with the students and helps them to develop this knowledge. Assessment for learning is incorporated into students', teachers', and peers' daily practice, which "seeks, reflects upon, and responds to information from dialogue, demonstration, and observation in ways that enhance ongoing learning” (Klenowski, 2009, p. 264). It is indicated in a study by Oyinloye and Imenda (2019) that the performance of the learners following an assessment for learning instructional approach outweighs those following normal classroom instruction.

### *Alternative Assessment*

Traditional assessment techniques gauge performance using objective questions with a single best or perfect response, usually in the form of a summative test (Brown & Abeywickrama, 2004). They are regarded as a one-shot, standardized, speed-based, and norm-referenced method of evaluating a behavior or performance that lacks authenticity (Bailey, 1998). Additionally, they are unable to determine a learner's progress and only assess what a student can achieve at a given point in time. Due to the aforementioned shortcomings, traditional forms of assessment have been replaced by alternative assessments. According to Smith (1999), alternative assessment encompasses various techniques employed continuously inside or outside the classroom to evaluate students’ knowledge in different ways and at different times. These evaluations are said to eventually lead to better instruction and open the door for

direct assessment of students' task achievement, using adaptable techniques (Kohonen, 1997). De-contextualized and artificial tasks and contextualized and realistic tasks are both used in assessment and each occupies one end of the assessment continuum. Modern evaluation techniques, however, incline toward the authentic end of the assessment spectrum in order to get students ready for the dynamic tasks found in everyday life (Boud, 1995). Studies indicate that alternative assessment methods are superior to traditional ones (Ahmad et al., 2020), and if applied properly, they can raise accomplishment since they measure the entire spectrum of student skills (Nasab, 2015). One main reason for this is that they seek to evaluate learners' communicative competence and not their decontextualized linguistic knowledge.

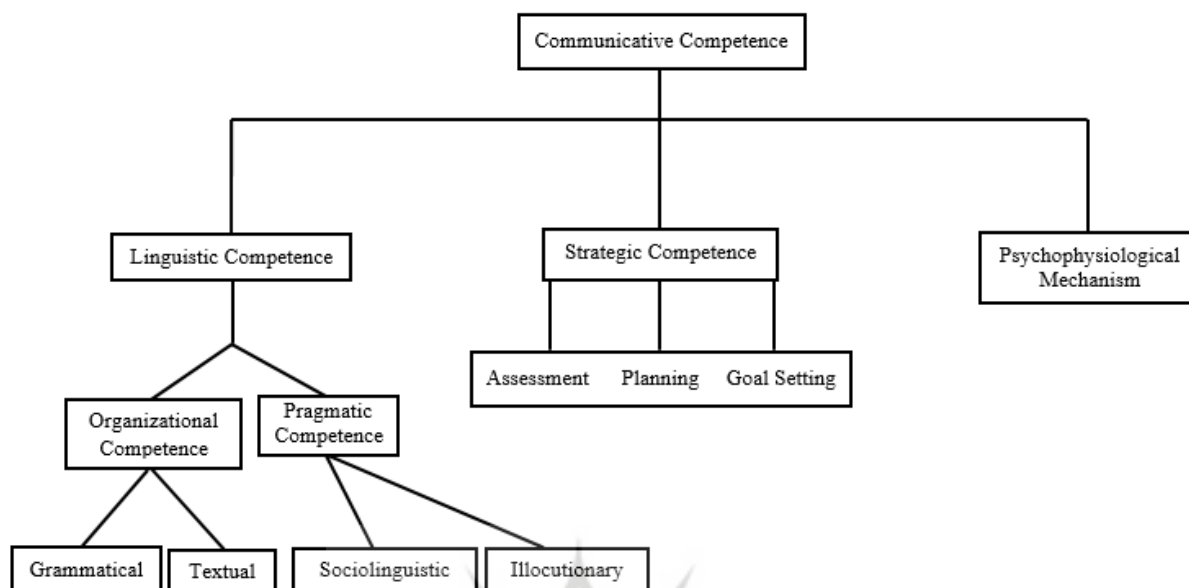
### *Communicative Competence*

In response to Chomsky's concept of linguistic competence, American sociolinguist and anthropologist, Dell H. Hymes, first used the term "communicative competence" in 1967. According to him, communicative competence is the capacity that "allows a member of the community to know which code to use, when, where, and to whom, etc." (Hymes, 1967, p. 13). The idea has evolved throughout time since then and different scholars have proposed several models of communicative competence. One of the models, which is dealt with in this study is Bachman and Palmer's (1996) model.

Bachman (1990) proposed a model of communicative competence known as "Communicative language ability" (CLA). Bachman and Palmer (1996) then made minor modifications to this model in the mid-1990s (Bagaric & Djigunovic, 2007). Three key elements comprise CLA: (1) language knowledge; (2) strategic competence; and (3) psychophysiological mechanisms. Organizational and pragmatic knowledge are the two main categories of language knowledge. Organizational knowledge regulates formal language structures in order to create or comprehend grammatically correct utterances or phrases (grammatical knowledge) and to arrange these utterances or sentences into written and spoken texts (textual knowledge) (Zohrabi & Jafari, 2020). Functional (illocutionary competence) and sociolinguistic knowledge are the two categories into which Bachman and Palmer (1996) subdivide pragmatic knowledge. By linking "utterances or sentences and texts to their meanings" as well as to the intentions of the language users, functional knowledge aids a person in comprehension of the discourse (Bachman & Palmer, 1996, p. 69). The four categories of language functions that make up functional knowledge are ideational,



manipulative, heuristic, and imaginative functions. Strategic competence also includes assessment, planning, and goal setting. Figure 1 illustrates Bachman and Palmer's (1996) model of communicative competence.



**Figure 1.** Bachman and Palmer's (1996) Model of Communicative Competence

### ***Communicative Language Testing***

The purpose of communicative language tests is to gauge language learners' capacity for engaging in conversation and interaction and making use of the language in everyday contexts. Communicative competence is the foundation for the construction of communicative tests, which address the four language skills of speaking, listening, reading, and writing. There are five prerequisites for developing a communicative test, including meaningful communication, authentic situations, unpredictable language input, creative language output, and integrated language skills (Brown, 2005). First and foremost, the test should be built around meaningful communication that fulfills the requirements of the learners. Then, it needs to encourage and stimulate language that is beneficial to them. Using real-life and contextualized situations can raise the possibility of meaningful communication because, according to Weir (1990, p. 11), "language cannot be meaningful if it is devoid of context." Moreover, in order to demonstrate students' level of language proficiency, communicative tests also provide them the opportunity to productively use the language in real-world contexts.

### ***Statement of the Problem***

Despite the changes in the theories of language teaching reflected in textbooks and principles of language testing, research shows that language tests still do not effectively assess students' communicative competence in the target language (Nguyen & Le, 2012). The test items are still artificial, fragmented, inauthentic, and unlikely to reflect language use in real-life situations. As Weir (1990) noted, "integrative tests such as cloze only tell us about a candidate's linguistic competence. They do not tell us anything directly about a student's performance ability" (p. 6). This situation is also evident in the context of Iran. Even though English textbooks have undergone changes in content and methodology by the Ministry of Education, this change cannot be appropriately witnessed in the content of the assessment tools. Considering this issue, the present study was conducted to investigate if the tests designed by the teachers reflect the recent changes in the updated version of textbooks or if they merely adhere to the traditional methods of assessing students' knowledge of the language without considering their communicative competence. It seeks answers to the following research questions:

1. To what extent do teacher-made tests match with the updated version of EFL textbooks in terms of content, format, and purpose?
2. What are the teachers' attitudes and insights toward the use of assessment tools alongside recent changes made by the Ministry of Education in EFL textbooks?

### **Methodology**

#### ***Participants***

Since the questionnaire was sent electronically through the Internet via different platforms including *Telegram*, *WhatsApp*, *Eitaa*, and *Instagram*, the sampling method employed was convenience sampling. Through this sampling method, individuals who were easily accessible voluntarily responded to the questionnaire sent to them through the mentioned platforms. The participants of the study were 30 (16 males and 14 females) English language teachers teaching tenth, eleventh, and twelfth grades at senior high schools in Iran. They were divided into three age groups: 20-29 (6 teachers), 30-39 (9 teachers), and 40-49 (15 teachers). Fourteen participants held an M.A. degree and 16 participants had a B.A. degree. Nine teachers had 10-20 years of teaching experience, 10 teachers had 5-9 years, 3 teachers had less than 5 years, and 8 teachers had more than 20 years of experience. Characteristics of the participants are presented in Table 1.

**Table 1.** Characteristics of the Participants

|                            |  |
|----------------------------|--|
| Number of the participants | 30   |
| Age range                  | 6 (20-29), 9 (30-39), 15 (40-49)                                   |
| Gender                     | 16 male, 14 female   |
| Degree                     | 14 (M.A degree), 16 (B.A degree)                                   |
| Experience                 | 9 (10-20), 10 (5-9), 3 (less than 5 years), 8 (more than 20 years) |

***Instruments***

A self-designed questionnaire was used in the present study to collect the quantitative data. It was adapted from related questionnaires to meet the research objectives. The questionnaire was prepared in English and evaluated for reliability and validity by getting help from English experts. A pilot study was conducted with a small group of participants to assess the questionnaire's clarity, comprehensibility, and internal consistency. Participants completed the questionnaire within a designated time frame, and their responses were analyzed using Cronbach's alpha. The results showed high internal consistency. Feedback from participants helped identify ambiguities and areas for improvement and minor modifications were made to enhance clarity. The successful completion of the pilot study and high internal consistency coefficient provide strong evidence for the questionnaire's reliability.

The questionnaire underwent rigorous validation to ensure its content and construct validity as well. Content validity was verified by a panel of experts who reviewed the items, ensuring they met research objectives and measured desired constructs. Exploratory and confirmatory factor analysis were employed to examine the construct validity of the questionnaire. The results showed a strong alignment between the questionnaire items and the theoretical framework, confirming its validity. The questionnaire was developed to measure teachers' knowledge about assessment techniques, their perceptions about the alignment between EFL textbooks and assessment methods, and factors influencing their practices.

In addition, the researchers developed a semi-structured interview protocol in order to gather qualitative data on the participants' attitudes toward the use of assessment tools. In order to develop this protocol, first, the researchers invited three professors of English Language Teaching at a university in Tabriz (Iran) to attend a focus-group interview session. Next, they prompted these professors to discuss the teachers' probable perspectives on the use of updated assessment tools and recorded the interview session. Lastly, the researchers transcribed the focus-group interview, used thematic analysis in order to extract its codes and themes, and developed the three-item semi-structured protocol of the study.



### ***Materials***

Materials used for the evaluation purpose in this study were a series of *Vision English for Schools* textbooks designed for grades 10, 11, and 12, and 100 teacher-made tests. The purpose of using these materials was to compare the content and the focus of these textbooks with the content and purpose of teacher-made tests to illustrate alignment and match between them.

### ***Design of the Study***

This study used an explanatory sequential mixed-methods design that combines quantitative data collected through questionnaires teacher-made tests, and textbooks with qualitative data gathered using a semi-structured interview protocol to provide a comprehensive understanding of the match between them.

### ***Data Collection Procedures***

The study was conducted in two stages. First, a descriptive qualitative approach was used to compare the content and focus of the updated English textbooks with the content, purpose, and format of the teacher-made tests to explore whether there is a match and alignment between them based on Brown and Abeywickrama's (2004) communicative language testing (CLT) framework. To this end, the researcher collected 100 teacher-made tests used in senior high school and compared them with the focus and the area of emphasis of *the Vision Series*.

Second, the researchers developed a questionnaire containing items that measure teachers' knowledge about assessment methods, their perceptions about the alignment between EFL textbooks and assessment methods, and the factors that influence their assessment practices. At first, a pilot study was conducted to see whether the collected data were useful and contributed to the main purpose of the present research. To this end, the researcher sent the questionnaire to a telegram channel whose members were English teachers, teaching to all three senior high school grades and asked them to participate in completing the questionnaire. After ten teachers completed the questionnaire, the researchers analyzed the collected data and found it helpful. After that, the number of answers to the questionnaire was increased to 30 by distributing it on different platforms like *Eitaa*, *Telegram*, *WhatsApp*, and *Instagram*.

Finally, the researchers conducted semi-structured interviews of the study to gather qualitative data on the teachers' assessment knowledge, perceptions of the compatibility of

the EFL textbooks and assessment practices, and factors in their language assessment. The interviews were conducted in Farsi and lasted about 30 minutes.

### ***Data Analysis***

With regard to the first research question, which aims to investigate the alignment between teacher-made tests and the updated EFL textbooks, the descriptive analysis was made by comparing the textbooks and tests based on Bachman and Palmer's (1996) model of communicative competence and Brown and Abeywickrama's (2004) classification of traditional assessment and alternative assessment. A self-designed questionnaire and a researcher-developed semi-structured interview protocol were also used to answer the second research question. The data were gathered by distributing the questionnaire on different online platforms and conducting the interviews. Finally, the data were analyzed using SPSS and thematic analysis. The statistical methods used to analyze the data were univariate chi-square. Moreover, the researchers transcribed the interviews and took advantage of thematic analysis in order to extract the codes and the themes in the interview data.

## **Results**

### ***Response to the First Research Question***

**Research Question 1:** To what extent do teacher-made tests match with the updated version of EFL textbooks in terms of content, format, and purpose?

### **Descriptive Analysis: Comparison of Vision Series with Teacher-Made Tests**

It was mentioned in the introduction section of all three versions of the Vision Series that the focus and the area of emphasis of these textbooks are to improve students' language skills so that they are able to communicate effectively in real-life situations. This is based on communicative language teaching theory which emphasizes improving students' communicative skills. The implemented changes in the content and purpose of the updated version of Vision Series are as mentioned below:

- Paying attention to the improvement of all four language skills simultaneously (listening, speaking, reading, writing);
- Using various authentic educational activities in the process of learning a language;
- Emphasizing learning the language through the use of language experiences;
- Using rich, meaningful, and comprehensible content in developing educational content;

- Providing opportunities for cooperative language learning;
- Providing learners with a positive washback;
- Paying attention to learners' affective factors in the process of teaching and learning.

As was stated above, there was a shift in the focus of textbooks from improving students' linguistic knowledge to the theory of communicative language teaching, which focuses on the enhancement of learners' communicative competence through presenting authentic language activities. By exploring the content of each lesson, it was confirmed that all four language skills including speaking, reading, writing, and speaking were equally emphasized through different tasks and exercises which encouraged learners to cooperate with each other and construct meaning jointly. However, regardless of these areas of emphasis in the textbooks, tests used by teachers dominantly relied on traditional ways of assessing students' language abilities and focused on test techniques such as multiple-choice questions, fill-in-the-blanks, true-false, matching, and essays. These testing techniques, according to Brown and Abeywickrama's (2004) classification of traditional and alternative assessment illustrated in Table 2, are regarded as traditional assessment methods, which focus on a definite "right" answer and cannot be used to assess students' interactive performance and therefore are not in line with the purpose of the updated English textbooks.

**Table 2.** Brown's classification of Traditional and Alternative Assessment

| Traditional assessment        | Alternative assessment             |
|-------------------------------|------------------------------------|
| One-shot, standardized exams  | Continuous long-term assessment    |
| Timed, multiple-choice format | Untimed, free-response format      |
| Decontextualized test items   | Contextualized communicative tasks |
| Scores suffice for feedback   | Individualized feedback & washback |
| Norm-referenced scores        | Criterion-referenced               |
| Focus on the "right" answer   | Open-ended, creative answers       |
| Summative                     | Formative                          |
| Oriented to product           | Oriented to process                |
| Non-interactive performance   | Interactive performance            |
| Fosters extrinsic motivation  | Fosters intrinsic motivation       |

**Note:** Adapted from Armstrong (1994) and Bailey (1998, p. 207)

The tests were also analyzed based on Bachman and Palmer's (1996) model of communicative competence. The components of language competence in this model are illustrated in Table 3. It is composed of two components: pragmatic and organizational

competence, which are further broken down into grammatical and textual competence and illocutionary and sociolinguistic competence, respectively. It was observed that test items are not designed in a way that they assess these competencies, except for grammatical ones. It is necessary to incorporate all aspects of the model, particularly pragmatic and strategic competence, into the constructions of language tests as well as the actual performance expected of the test-takers. In general, it can be said that implementing traditional assessment methods alongside the use of newly-developed textbooks, the aim of which is to teach a communicative way of learning a language and improve students' communicative skills, does not lead to a consistent way of teaching a language.

**Table 3.** Bachman and Palmer's (1996) Model of Communicative Competence

| Components of language competence  |  |
|--|--|
| A, Organizational Competence   |  |
| 1. Grammatical (including lexicon, morphology, and phonology)            |  |
| 2. Textual (discourse)   |  |
| B. Pragmatic Competence  |  |
| 1. Illocutionary (functions of language)                                 |  |
| 2. Sociolinguistic (including culture, context, pragmatics, and purpose) |  |

**Note:** Adapted from Bachman (1990, p. 87)

### Frequency Analysis

The frequency of opinions related to the rate of match of the tests and frequency of the opinions related to the mismatch of the tests were also examined through Chi-square. The results are presented in the following table:

**Table 4.** Results of Chi-square Test to Examine Frequency Differences

| Question  | Component         | Observed N | Expected N | Residual | Chi-Square | df | Sig.  |
|---|-------------------|------------|------------|----------|------------|----|-------|
| Is there a match between the theory of your tests and the theory that the new version textbooks are based on? | Yes               | 9          | 15.0       | -6.0     | 4.800a     | 1  | .028  |
|   | No                | 21         | 15.0       | 6.0      |            |    |       |
|   | Total             | 30         |            |          |            |    |       |
| There is a need to change assessment methods alongside the changes in methodology and textbooks.              | Strongly disagree | 1          | 10.0       | -9.0     | 16.20      | 2  | 0.001 |
|   | Agree             | 10         | 10.0       | .0       |            |    |       |
|   | Strongly agree    | 19         | 10.0       | 9.0      |            |    |       |
| Teaching and testing are inseparable, so every change in  | Strongly disagree | 1          | 7.5        | -6.5     | 32.13      | 3  | 0.001 |

|  |                   |    |     |      |      |   |       |  |
|--|-------------------|----|-----|------|------|---|-------|--|
| methodology requires changes in assessment methods too.                                      | disagree          | 1  | 7.5 | -6.5 |      |   |       |  |
|  | Agree             | 8  | 7.5 | .5   |      |   |       |  |
|  | Strongly agree    | 20 | 7.5 | 12.5 |      |   |       |  |
|  | Total             | 30 |     |      |      |   |       |  |
| The tests that we use in our classes violate the purpose and focus of new version textbooks. | Strongly disagree | 1  | 7.5 | -6.5 | 8.93 | 3 | 0.030 |  |
|  | disagree          | 7  | 7.5 | -.5  |      |   |       |  |
|  | Agree             | 11 | 7.5 | 3.5  |      |   |       |  |
|  | Strongly agree    | 11 | 7.5 | 3.5  |      |   |       |  |
|  | Total             | 30 |     |      |      |   |       |  |

The results of the chi-square test in the above table show that there is a significant difference between the frequency of the teachers' opinions, who believe that there is no match between the theory of their tests and the theory that the new version of textbooks is based on, with the frequency of the teachers' opinions who believe otherwise. The value of the chi-square test is 4.80 and its level of significance is 0.028 ( $<0.05$ ). Therefore, it could be indicated that there is no match between the theory of teacher-made tests and that of the textbooks.

In the second item, the value of the chi-square test is 16.20 and its level of significance is 0.01 ( $< 0.05$ ), which indicates that the teachers believe that besides the existing changes in methodology and textbooks, there should also be changes in the assessment principles. Furthermore, regarding the third item, the value of the chi-square test is significant and the participants believe that alongside changes in methodology, change in assessment methods is necessary and there should be a match between them. In item four, the value of the chi-square test is significant at the level of 0.05 and it demonstrates that most of the teachers believe that the tests they use in their classes, do not match the objectives of the new textbooks.

### ***Response to the Second Research Question***

**Research Question 2:** What are the teachers' attitudes and insights toward the use of assessment tools alongside recent changes made by the Ministry of Education in EFL textbooks?

### **Quantitative Analysis**

To answer this question, the univariate chi-square test was employed and the observed frequencies in several ranks with the expected frequencies were examined. In the following table, results of the chi-square test examining the differences of the observed frequencies



with regard to the rate of teachers' knowledge of the components of the test are presented. The answers for the following three questions were classified as 'not at all' with code 0, 'overview or introduction to topic' with code 1, and 'it was an area of emphasis' with code 2.

**Table 5.** Results of Chi-square Test to Answer Related Questions to the Rate of Teachers' Study toward the Test Aspects

As part of your formal education and/or training, to what extent did you study the following areas?

|  |       | Observed<br>N | Expected<br>N | Residual | Chi-<br>Square | df | Sig.<br>(2-tailed) |
|--|-------|---------------|---------------|----------|----------------|----|--------------------|
| Test language                                  | 0     | 2             | 10.0          | -8.0     | 12.80          | 2  | 0.002              |
|  | 1     | 18            | 10.0          | 8.0      |                |    |                    |
|  | 2     | 10            | 10.0          | .0       |                |    |                    |
|  | Total | 30            |               |          |                |    |                    |
| Pedagogy/Teaching                              | 0     | 1             | 10.0          | -9.0     | 16.20          | 2  | 0.001              |
|  | 1     | 10            | 10.0          | .0       |                |    |                    |
|  | 2     | 19            | 10.0          | 9.0      |                |    |                    |
|  | Total | 30            |               |          |                |    |                    |
| Theoretical models and<br>processes of testing | 0     | 2             | 10.0          | -8.0     | 19.40          | 2  | 0.001              |
|  | 1     | 21            | 10.0          | 11.0     |                |    |                    |
|  | 2     | 7             | 10.0          | -3.0     |                |    |                    |
|  | Total | 30            |               |          |                |    |                    |

The results of the above table indicate that the frequency of teachers' responses regarding the three areas of 'test language', 'pedagogy/teaching', and 'theoretical models and processes of testing', is statistically significant. A comparison of the frequencies also shows that the frequencies of response options 1 and 2 are higher than option 0.

In the following table, the results of the chi-square test examining the frequencies of the teachers' attitudes toward the effect of changes on four language skills are demonstrated. The answers are represented in a four-point Likert scale including 'very improved', 'improved', 'somewhat improved', and 'neutral', ranging from 0 to 3, respectively.

**Table 6.** The Results of Chi-square for the Items Related to Teachers' Attitude toward the Effect of Changes on Fourfold Language Skills

To what extent do the changes in the content of textbooks and methodology improve students' language skills, including

|            |       | Observed N | Expected N | Residual | Chi-Square          | Df | Sig.<br>(2-tailed) |
|------------|-------|------------|------------|----------|---------------------|----|--------------------|
| Speaking?  | 0     | 1          | 7.5        | -6.5     | 33.200 <sup>a</sup> | 3  | .000               |
|            | 1     | 4          | 7.5        | -3.5     |                     |    |                    |
|            | 2     | 4          | 7.5        | -3.5     |                     |    |                    |
|            | 3     | 21         | 7.5        | 13.5     |                     |    |                    |
|            | Total | 30         |            |          |                     |    |                    |
| Listening? | 0     | 2          | 7.5        | -5.5     | 16.133 <sup>a</sup> | 3  | .001               |
|            | 1     | 2          | 7.5        | -5.5     |                     |    |                    |
|            | 2     | 13         | 7.5        | 5.5      |                     |    |                    |
|            | 3     | 13         | 7.5        | 5.5      |                     |    |                    |
|            | Total | 30         |            |          |                     |    |                    |
| Reading?   | 0     | 2          | 7.5        | -5.5     | 21.467 <sup>a</sup> | 3  | .000               |
|            | 1     | 18         | 7.5        | 10.5     |                     |    |                    |
|            | 2     | 7          | 7.5        | -.5      |                     |    |                    |
|            | 3     | 3          | 7.5        | -4.5     |                     |    |                    |
|            | Total | 30         |            |          |                     |    |                    |
| Writing?   | 0     | 1          | 7.5        | -6.5     | 7.867 <sup>a</sup>  | 3  | .049               |
|            | 1     | 11         | 7.5        | 3.5      |                     |    |                    |
|            | 2     | 9          | 7.5        | 1.5      |                     |    |                    |
|            | 3     | 9          | 7.5        | 1.5      |                     |    |                    |
|            | Total | 30         |            |          |                     |    |                    |

Based on the results, it is observed that according to the teachers' opinions, changes in the textbooks' content will not lead to improved speaking, listening, and writing skills. However, with regard to reading, 18 teachers believed that changing the content of the textbooks, to some extent would lead to the improvement of the students' reading skills.

Table 7 shows the frequency and the results of the chi-square test related to the teachers' attitudes toward assessment. The answers are represented in a four-point Likert

scale, including 'strongly disagree', 'disagree', 'agree', and 'strongly agree', ranging from 0 to 3, respectively.

**Table 7.** The Results of Chi-square for the Items Related to the Teachers' Attitudes toward Assessment

| How much do you agree with the following statements?  |               |               |          |                |                     |                     |
|---|---------------|---------------|----------|----------------|---------------------|---------------------|
|   | Observed<br>N | Expected<br>N | Residual | Chi-<br>Square | Df                  | Sig. (2-<br>tailed) |
|   | 0             | 1             | 10.0     | -9.0           |                     |                     |
| There is a need to change assessment methods alongside The changes in methodology and textbooks.  | 1             | 0             | 0        | 0              | 16.200 <sup>a</sup> | 2                   |
|   | 2             | 10            | 10.0     | .0             |                     |                     |
|   | 3             | 19            | 10.0     | 9.0            |                     |                     |
|   | Total         | 30            |          |                |                     |                     |
| Teaching and testing are inseparable, so every change in methodology requires changes in assessment methods too.  | 1             | 7.5           | -6.5     |                | 32.133 <sup>b</sup> | 3                   |
|   | 1             | 7.5           | -6.5     |                |                     |                     |
|   | 8             | 7.5           | .5       |                |                     |                     |
|   | 20            | 7.5           | 12.5     |                |                     |                     |
|   | 30            |               |          |                |                     |                     |
| The tests that we use in our classes violate the purpose and focus of new version textbooks.  | 0             | 1             | 7.5      | -6.5           | 8.933 <sup>b</sup>  | 3                   |
|   | 1             | 7             | 7.5      | -.5            |                     |                     |
|   | 2             | 11            | 7.5      | 3.5            |                     |                     |
|   | 3             | 11            | 7.5      | 3.5            |                     |                     |
| Total   | 30            |               |          |                |                     |                     |
| The tests that we use assess only what students have learned (their linguistic knowledge including grammar, vocabulary, ...)  | 0             | 0             | 0        | 0              | 7.400 <sup>a</sup>  | 2                   |
|   | 1             | 3             | 10.0     | -7.0           |                     |                     |
|   | 2             | 13            | 10.0     | 3.0            |                     |                     |
|   | 3             | 14            | 10.0     | 4.0            |                     |                     |
|   | Total         | 30            |          |                |                     |                     |
| Tests must be designed in a way that not only evaluate students' linguistic knowledge but also evaluate their communicative skills and at the same time enhance their learning and make them ready to communicate effectively in real life. | -             | -             | -        | -              | 4.800 <sup>c</sup>  | 1                   |
|   | -             | -             | -        | -              |                     |                     |
|   | 2             | 9             | 15.0     | -6.0           |                     |                     |
|   | 3             | 21            | 15.0     | 6.0            |                     |                     |
|   | Total         | 30            |          |                |                     |                     |
| The tests that we use are in line with the purpose of new textbooks and these tests enhance students' learning  | 0             | 9             | 7.5      | 1.5            | 4.667 <sup>b</sup>  | 3                   |
|   | 1             | 11            | 7.5      | 3.5            |                     |                     |
|   | 2             | 7             | 7.5      | -.5            |                     |                     |

|   |       |    |     |      |
|---|-------|----|-----|------|
| and make them ready to transfer their knowledge to real life. | 3     | 3  | 7.5 | -4.5 |
|   | Total | 30 |     |      |

According to the results, except for the last item, in the remaining items, there is no significant difference in the observed frequencies, and the results of the chi-square test are significant and the two options of 'agree' and 'strongly agree' are more frequently used.

The following table indicates the frequency of the teachers' responses regarding the role of assessment in improving communicative skills. The responses are represented in a four-point Likert scale, including 'not important', 'somewhat important', 'important', and 'very important', ranging from 0 to 3, respectively.

**Table 8.** The Results of Chi-square for the Items Related to the Teachers' Attitude toward the Role of Assessment in Improved Communicational Skills

|   |       | Observed<br>N | Expected<br>N | Residual | Chi-<br>Square      | df | Sig. (2-<br>tailed) |
|---|-------|---------------|---------------|----------|---------------------|----|---------------------|
| How important do you consider the role of changing assessment tools and using alternative assessments instead of traditional assessments, alongside the changes in textbooks and methodology in improving students' communicative skills? | 0     | 1             | 7.5           | -6.5     |                     |    |                     |
|   | 1     | 3             | 7.5           | -4.5     |                     |    |                     |
|   | 2     | 11            | 7.5           | 3.5      |                     |    |                     |
|   | 3     | 15            | 7.5           | 7.5      | 17.467 <sup>a</sup> | 3  | .001                |
|   | Total | 30            |               |          |                     |    |                     |

Results indicate that the value of the chi-square test is significant at the level of 0.001 and options 'important' and 'very important' are used more frequently. Therefore, teachers perceive changing assessment tools as important and believe it to be necessary in developing students' communicative skills.

The table below shows the frequencies of the teachers' responses about how much satisfied they are with their knowledge about several fields of assessment. The answers are represented in a four-point Likert scale, including 'very dissatisfied', 'dissatisfied', 'satisfied', and 'very satisfied', ranging from 0 to 3, respectively.

**Table 9.** The Results of Chi-square for the Items Related to Teachers' Attitude toward Satisfaction of Their Knowledge about Assessment

Please look at the following language testing and assessment-related topics, and rate your level of satisfaction with your knowledge of them

|  |       | Observed<br>N | Expected<br>N | Residual | Chi-Square          | df | Sig. (2-tailed) |
|--|-------|---------------|---------------|----------|---------------------|----|-----------------|
| History of language<br>testing   | 0     | 2             | 7.5           | -5.5     | 13.467 <sup>a</sup> | 3  | .004            |
|  | 1     | 4             | 7.5           | -3.5     |                     |    |                 |
|  | 2     | 9             | 7.5           | 1.5      |                     |    |                 |
|  | 3     | 15            | 7.5           | 7.5      |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |
| Design of language<br>assessments for speaking,<br>listening, reading, writing   | 0     | 1             | 7.5           | -6.5     | 9.200 <sup>a</sup>  | 3  | .027            |
|  | 1     | 7             | 7.5           | -.5      |                     |    |                 |
|  | 2     | 10            | 7.5           | 2.5      |                     |    |                 |
|  | 3     | 12            | 7.5           | 4.5      |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |
| Deciding what to test,<br>writing test specifications<br>/writing test tasks and<br>items  | -     | -             | -             | -        | 18.600 <sup>b</sup> | 2  | .000            |
|  | 1     | 3             | 10.0          | -7.0     |                     |    |                 |
|  | 2     | 6             | 10.0          | -4.0     |                     |    |                 |
|  | 3     | 21            | 10.0          | 11.0     |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |
| Interpreting and analyzing<br>test scores  | -     | -             | -             | -        | 16.800 <sup>b</sup> | 2  | .000            |
|  | 1     | 2             | 10.0          | -8.0     |                     |    |                 |
|  | 2     | 8             | 10.0          | -2.0     |                     |    |                 |
|  | 3     | 20            | 10.0          | 10.0     |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |
| Reliability of tests<br>/validity of tests   | 0     | 1             | 7.5           | -6.5     | 16.667 <sup>a</sup> | 3  | .001            |
|  | 1     | 3             | 7.5           | -4.5     |                     |    |                 |
|  | 2     | 14            | 7.5           | 6.5      |                     |    |                 |
|  | 3     | 12            | 7.5           | 4.5      |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |
| Authenticity in language<br>assessment real-life tasks<br>communicative language<br>testing (CLT) task-based<br>assessment (TBA) | 0     | 2             | 7.5           | -5.5     | 9.200 <sup>a</sup>  | 3  | .027            |
|  | 1     | 12            | 7.5           | 4.5      |                     |    |                 |
|  | 2     | 5             | 7.5           | -2.5     |                     |    |                 |
|  | 3     | 11            | 7.5           | 3.5      |                     |    |                 |
|  | Total | 30            |               |          |                     |    |                 |



|   |       |    |      |      |                     |   |      |
|---|-------|----|------|------|---------------------|---|------|
|   | -     | -  | -    | -    |                     |   |      |
| Scoring closed-response items scoring open-response test tasks          | 1     | 2  | 10.0 | -8.0 |                     |   |      |
|   | 2     | 14 | 10.0 | 4.0  | 9.600 <sup>b</sup>  | 2 | .008 |
|   | 3     | 14 | 10.0 | 4.0  |                     |   |      |
|   | Total | 30 |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
| Test-taking skills or strategies /test administration and accommodation | 0     | 1  | 7.5  | -6.5 |                     |   |      |
|   | 1     | 6  | 7.5  | -1.5 |                     |   |      |
|   | 2     | 13 | 7.5  | 5.5  | 10.800 <sup>a</sup> | 3 | .013 |
|   | 3     | 10 | 7.5  | 2.5  |                     |   |      |
|   | Total | 30 |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
| The use of tests in the society   | 0     | 3  | 7.5  | -4.5 |                     |   |      |
|   | 1     | 9  | 7.5  | 1.5  |                     |   |      |
|   | 2     | 10 | 7.5  | 2.5  | 3.867 <sup>a</sup>  | 3 | .276 |
|   | 3     | 8  | 7.5  | .5   |                     |   |      |
|   | Total | 30 |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
| Norm-referenced vs. criterion-referenced testing                        | 0     | 1  | 7.5  | -6.5 |                     |   |      |
|   | 1     | 5  | 7.5  | -2.5 | 12.133 <sup>a</sup> | 3 | .007 |
|   | 2     | 13 | 7.5  | 5.5  |                     |   |      |
|   | 3     | 11 | 7.5  | 3.5  |                     |   |      |
|   | Total | 30 |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |
| Washback on the classroom   | 1     | 1  | 10.0 | -9.0 |                     |   |      |
|   | 2     | 7  | 10.0 | -3.0 | 23.400 <sup>b</sup> | 2 | .000 |
|   | 3     | 22 | 10.0 | 12.0 |                     |   |      |
|   | Total | 30 |      |      |                     |   |      |
|   |       |    |      |      |                     |   |      |

Based on the results, except for item "the use of tests in society", there is a significant difference in the observed frequencies of the responses, and options 2 and 3 were mostly frequent, indicating that the rate of teachers' satisfaction with their own knowledge about assessment is optimal.

Frequencies of the teachers' responses about general perception of their own knowledge about assessment are represented in the following table. The responses to this item are represented in a four-point Likert scale, including 'very prepared', 'somewhat prepared', 'somewhat unprepared', and 'very unprepared', ranging from 0 to 3, respectively.

**Table 10.** The Results of Chi-square for the Items Related to the Teachers' Attitude toward the General Perception of their own Knowledge about Assessment

|                        | Observed N | Expected N | Residual | Chi-Square | df                  | Sig. (2-tailed) |
|------------------------|------------|------------|----------|------------|---------------------|-----------------|
| Which of the following | 0          | 4          | 7.5      | -3.5       | 13.200 <sup>a</sup> | 3               |
| best describes your    | 1          | 16         | 7.5      | 8.5        |                     |                 |
| perception of your     | 2          | 6          | 7.5      | -1.5       |                     |                 |
| overall knowledge and  | 3          | 4          | 7.5      | -3.5       |                     |                 |
| understanding of       | Total      | 30         |          |            |                     |                 |
| language assessment?   |            |            |          |            |                     | .004            |

According to the results, the value of the chi-square test is significant and it shows that there is a significant difference in the observed frequencies of the response options and 16 teachers choose the option 'somewhat prepared', meaning that they generally consider their knowledge of language assessment somehow acceptable.

Finally, the following table demonstrates the frequencies of the teachers' responses to the extent of their familiarity with several assessment approaches. The responses are represented in a three-point Likert scale, including 'not at all', 'somewhat familiar', and 'familiar', ranging from 0 to 2, respectively.

**Table 11.** The Results of Chi-square for the Items Related to the Teachers' Rate of Familiarity with Several Assessment Approaches

| To what extent are you familiar with different approaches of testing? |            |            |          |            |                     |                 |
|---|------------|------------|----------|------------|---------------------|-----------------|
|   | Observed N | Expected N | Residual | Chi-Square | df                  | Sig. (2-tailed) |
| Discrete-point testing  | 0          | 3          | 10.0     | -7.0       | 11.400 <sup>a</sup> | 2               |
|   | 1          | 9          | 10.0     | -1.0       |                     |                 |
|   | 2          | 18         | 10.0     | 8.0        |                     |                 |
|   | Total      | 30         |          |            |                     |                 |
| Integrative testing   | 0          | 1          | 10.0     | -9.0       | 12.600 <sup>a</sup> | 2               |
|   | 1          | 13         | 10.0     | 3.0        |                     |                 |
|   | 2          | 16         | 10.0     | 6.0        |                     |                 |
|   | Total      | 30         |          |            |                     |                 |
| Communicative language testing (CLT)                                  | 0          | 2          | 10.0     | -8.0       | 9.600 <sup>a</sup>  | 2               |
|   | 1          | 14         | 10.0     | 4.0        |                     |                 |
|   | 2          | 14         | 10.0     | 4.0        |                     |                 |
|   | Total      | 30         |          |            |                     |                 |

The results of chi-square are significant with regard to all three approaches of assessment and the examination of frequencies of the response options shows that most of the teachers believe that they are familiar with different approaches.

### Qualitative Analysis

In addition to the analysis of the questionnaire data, the researchers examined the interview data to delve into the participants' attitudes towards the assessment tools. The thematic analysis of the collected data highlighted the existence of three underlying themes. Table 12 shows these themes along with their pertinent codes:

**Table 12.** Codes and Themes in the Interview Data

| Codes   | Themes  |
|---|---|
| Studying various sources  |   |
| Watching language assessment videos   | Satisfactory assessment knowledge   |
| Making an endeavor to use learning-oriented assessment practices                    |   |
| Disregarding the significance of the assessment of all language skills              | Lack of compatibility between the content of EFL textbooks and modern assessment approaches |
| Developing materials without paying attention to the integration of language skills |   |
| Focusing on receptive skills  |   |
| Emphasizing the need to revise EFL textbooks based on modern assessment approaches  |   |
| Highlighting the utility of alternative assessment techniques                       | Openness to language-assessment-oriented changes  |
| Underlining the need to integrate language assessment into language learning        |   |

As shown in Table 12, the first theme in the obtained interview data was *satisfactory assessment knowledge*. Twenty-two of the participants stated that they had adequate information on the diverse aspects of language assessment. In this regard, participant 4 stated that:

“I have studied most of the commercial textbooks of language assessment. In fact, I was interested in this area of language instruction during my university

studies and made an effort to translate the theory of language assessment into practice in my general English classes”.

Likewise, participant 23 highlighted the fact that her interest in language assessment prompted her to watch numerous videos on YouTube and Instagram. As she explained:

“Reading books and articles may not be enough. I want to emphasize the fact that they might not help me to develop a practical knowledge of language assessment. As a result, I watch language assessment videos on YouTube and some Instagram pages. These applications have made me aware that learning-oriented assessment has become the most popular language assessment approach in diverse contexts”.

Moreover, according to Table 12, the second major theme in these participants’ data was *the lack of compatibility between the content of EFL textbooks and modern assessment approaches*. Nineteen of the participants stated that the EFL textbooks were developed in disregard of the current theoretical discussions of language assessment. For instance, participant 18 noted that:

“Our textbooks follow the traditional language testing procedures. The examination of their content shows that they ignore the assessment of some skills such as writing. Moreover, they emphasize the individual assessment of skills without paying attention to their integration in communication”.

Similarly, participant 26 pointed out that:

“What are the modern language assessment approaches? Scholars talk about dynamic assessment and learning-oriented assessment. What about our textbooks? They are developed based on the traditional language teaching methods and consider reading to be the most important language skill which must be practiced by using various exercises”.

Lastly, as shown in Table 12, the third theme in these participants’ interview data was *openness to language-assessment-oriented changes*. Twenty-four of the participants underlined the usefulness of innovations in the field of language assessment. In this regard, participant 11 noted that:

“It is necessary to change these textbooks thoroughly. Their content reminds me of the Grammar Translation Method. We must use new assessment approaches such as portfolio assessment which focus on the process of language learning and provide a better understanding of the learners’ strengths and weaknesses”.

Likewise, participant 29 underlined the importance of learning-oriented assessment in language classes. As he explained:

“It has become clear that assessment constitutes a major aspect of learning the language. That is, it cannot be distinguished from the process of learning. As a result, we need to revise our textbooks to take advantage of assessment for improving the learners’ acquisition of the target language”.

## Discussion

The present research was carried out in a quantitative phase and a qualitative phase. The quantitative phase was conducted to compare the purpose and the theory behind EFL textbooks (here Vision series) and teacher-made tests based on Bachman and Palmer’s (1996) model of communicative competence and Brown and Abeywickrama's (2004) classification of traditional assessment and alternative assessment. The results indicated that even though there was a shift away from emphasis on improving learners' linguistic competence, and toward enhancing their communicative skills in the newly-designed textbooks, this was not reflected in the theory and content of teacher-made tests. In other words, the tests were not designed by teachers in a way that they elicit the competencies that were intended to be enhanced by EFL textbooks.

Moreover, the quantitative phase was conducted through questionnaires to provide answers to the two research questions. First, the results regarding the first research question, which was posed to explore the extent to which teachers believe that the content and objective of their designed tests align with the content and objectives determined for the newly-developed EFL textbooks, indicate that they don't perceive any match between the theory of their tests and the theory on which the newly-designed EFL textbooks are based. This finding is somehow consistent with Riazi and Mosalanejad's (2010) declaration that textbooks are unable to meet the different needs and demands of those who use them and are inherently shallow and reductionist. In addition, Fidian and Supriani (2018) also demonstrated in their study that the syllabus's fundamental competencies are not met by the content found in the textbooks. Their investigation also revealed that the content of textbooks mostly does not adhere to the core competencies listed in the English Curriculum 2013. Besides, some of the books do not satisfy the demands and desires of the students either. Furthermore, the results showed that according to teachers' opinions, the change in language pedagogy not only should take place in methodology but also should be evident in methods of assessment. In other words, emphasizing the development of communicative competence in teaching calls for assessment focusing on evaluating this competence as well.



The second research question was posed to unveil teachers' attitudes and opinions toward the use of assessment tools and techniques that conform to the recent changes made by the Ministry of Education in EFL textbooks. First of all, the results of the chi-square showed that the teachers' study rate of different aspects of testing was at a satisfactory level. It was also demonstrated that according to teachers' opinion, changes in the content of textbooks will not lead to improvement of speaking, listening, and writing skills, but improve reading skills to some extent. This might be due to the fact EFL textbooks mostly include passages that are designed for developing students' reading comprehension and other supplementary activities might not suffice for the improvement of other skills. They also perceived changing assessment tools and altering new ones as crucial and believed that these assessment tools are necessary for developing students' communicative skills.

Moreover, results concerning the teachers' perception of their knowledge about language testing and assessment-related topics indicated that they believed their knowledge to be at a satisfactory and optimal level, except for the use of tests in society. In addition, teachers were also demonstrated to be familiar with most of the approaches of testing and considered their knowledge of this field acceptable. Lastly, the qualitative phase examined the participants' attitudes towards assessment tools using a semi-structured interview protocol. The findings of qualitative thematic analysis in this phase of data collection corroborated the quantitative frequency analysis. More specifically, they showed that language teachers had adequate knowledge about language assessment, were not satisfied with the content of the EFL textbooks in terms of their assessment practices, and were open to language-assessment-based changes in their field.

## Conclusion

The focus of language teaching has long shifted from the development of grammar and vocabulary to improving learners' all four language skills, to help them transfer their knowledge of language to real-life situations. This change has also been reflected in EFL textbooks. Foreign language textbooks can help students enhance their language proficiency and intercultural communication skills by not only providing target forms and meanings as language input but also by providing cultural and ideological inputs and perhaps modeling learning practices (Xiang & Yenika-Agbaw, 2021). Iranian schools use resources and materials created in accordance with the curriculum and syllabuses developed by the Ministry of Education. After nearly a quarter of a century, the Ministry of Education in Iran began to alter the junior and senior high school English textbooks since these materials have always

been the focus of intense discussion and criticism regarding their overall content, technical excellence, and applicability. A frequently asked question is why after seven years of English instruction in our schools, graduated students can't still communicate in English properly and convey simple messages. One answer to this question is that for developing students' communicative competence, changes in the content of textbooks and methodology do not suffice, and assessment methods need to be improved as well.

The current conditions in Iranian secondary schools also highlight the need for administering changes not only in methodology and teaching but also in methods of evaluation because teaching and testing are two integrated processes. Hence, using alternative assessment tools that are in line with the focus and purpose of the updated version of textbooks to satisfy the needs and requirements of language programs is of paramount importance. Considering this issue, an attempt was made in the present study to demonstrate whether teachers have changed their assessment methods alongside the recent changes in textbooks or they solely rely on traditional ways of assessing students' knowledge using paper and pencil tests which emphasize their linguistic competence regardless of their communicative capacity (Zohrabi & Tahmasebi, 2020). The results showed that teachers do not believe that there is a match between the theory of their tests and that of the newly-designed EFL textbooks, and confirmed that in order to develop students' communicative skills, changing assessment tools and altering new ones are crucial.

The findings of the present study might have some important pedagogical implications both for teachers and students, urging them to redefine their responsibilities. In a broad sense, this study can help teachers to remember that changes in the content of textbooks and methodology are not enough to improve students' communicative skills and they should implement some changes in assessment methods as well as design tests that elicit behaviors required in authentic situations. Moreover, the results may have two practical implications for the syllabus designers. First, syllabus designers have to revise the content of the current EFL textbooks in light of the modern approaches to language assessment including learning-oriented assessment. The inclusion of the learning-oriented assessment tasks in these textbooks may enable language teachers to use the assessment procedures for improving the learners' acquisition of the diverse aspects of the target language. Second, the syllabus designers need to revise the teacher manuals. More specifically, they need to include a specific section in these manuals which provides the teachers with adequate information

about the modern language assessment approaches and empowers them to translate their theory into practice in the context of the classroom.

As with any research, the current study fell prey to several limitations as well. First of all, the instrument used was only a questionnaire, and the data were gathered online, so the number of participants was not high enough to make valid generalizations about the findings. Furthermore, the participants of the study were only teachers, and involving students as important actors is crucial for obtaining more valid and reliable results.

## References

- Ahmad, S., Sultana, N., & Jamil, S. (2020). Behaviorism vs constructivism: A paradigm shifts from traditional to alternative assessment techniques. *Journal of Applied Linguistics and Language Research*, 7(2), 19-33.
- Armstrong, T. (1994). Multiple intelligences: Seven ways to approach curriculum. *Educational Leadership*, 52(3), 26-28.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. New York: Oxford University Press.
- Bachman, L. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Bagaric, V., & Djigunovic, J. M. (2007). Defining communicative competence. *Metodika*, 8(1), 94-103.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Pacific Grove, CA: Heinle & Heinle.
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers?. *ELT Journal*, 73(2), 113–123.
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98.
- Boud, D. (1995). Assessment and learning: contradictory or complementary? In: Knight, P. (Ed.) *Assessment for learning in Higher Education* (London, Kogan Page).
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In Liamputtong, P. (Ed.), *Handbook of research methods in health social sciences* (pp. 843–860). Springer.

- Brown, H. D., & Abeywickrama, P. (2004). *Language assessment; Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill.
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455.
- Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education*, 29(10), 1261-1280.
- Fidian, A., & Supriani, N. (2018). Conformity of English textbooks for seventh grade with curriculum 2013. *Journal of English Language and Pedagogy*, 1(1).
- Fulcher, G. (2020). Operationalizing language assessment literacy. In Tsagari, D. (Ed.), *Language assessment literacy: Theory to practice* (pp.7-28). Cambridge Scholar Publishing.
- Gilani, S. A., Ismail, N. F. B. M., Kassim, R. R. B. M., Yawen, J., & Dan, M. (2021). A comprehensive analysis of research on dynamic assessment in EFL speaking context. *AJELP: Asian Journal of English Language and Pedagogy*, 9(1), 65-79.
- Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 22(1), 189–200.
- Giraldo, F. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 23(1), 197–213.
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual review of applied linguistics*, 15, 212-226.
- Harding, L., & Brunfaut, T. (2020). Trajectories of language assessment literacy in a teacher-researcher partnership: Locating elements of praxis through narrative inquiry. In: Poehner, M. E. and Inbar-Lourie, O. (Eds.), *Towards a re-conceptualization of second language classroom assessment* (pp. 61–81). Springer International Publishing.
- Hymes, H. D. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, 23(2), 8-28.

- Isbell, D. R., Kremmel, B., & Kim, J. (2023). Remote proctoring in language testing: Implications for validity, fairness, and justice. *Language Assessment Quarterly*, 20(5), 469–487.
- Jiang, Y. C., Jong, S. Y., Lau, W. F., Chai, C. S., Liu, S. X., & Park, M. Y. (2022). A scoping review on flipped classroom approach in language education: Challenges, implications and an interaction model. *Computer Assisted Language Learning*, 35(6), 1218–1249.
- Kim, A. A., Chapman, M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy required for interpreting score reports: A focus on educators of K–12 English learners. *Language Testing*, 37(1), 54–75.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 263–268.
- Kohonen, V. (1997). Authentic Assessment as an Integration of Language Learning, Teaching, Evaluation and the Teacher's Professional Growth. In: Huhta, A., Kohonen, V., Kurki-Suonio, L., and Luoma, S. (eds.), *Current development and alternatives in language assessment*. University of Jyväskylä: Jyväskylä. Pp.7- 22.
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120.
- Kulieke, M., Bakker, J., Collins, C., Fennimore, T., Fine, C., Herman, J., Jones, B. F., Raack, L., & Tinzmann, M. B. (1990). *Why should assessment be based on a vision of learning?* [online document] NCREL, Oak Brook: IL. Available online: [http://www.ncrel.org/sdrs/areas/rpl\\_esys/assess.htm](http://www.ncrel.org/sdrs/areas/rpl_esys/assess.htm)
- Lee, J. Y., & Butler, Y. (2020). Reconceptualizing language assessment literacy: Where are language learners?. *TESOL Quarterly*, 54(4), 1098–1111.
- Lee, J., Butler, Y. G., & Peng, X. (2021). Multiple stakeholder interaction to enhance pre-service teachers' language assessment literacy. *Languages*, 6(4), 213.
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182.
- Marzano, R. J. (2006). *Classroom assessment & grading that work*. ASCD.
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Nguyen, C., & Le, D. (2012). Communicative language testing: Do school tests measure students' communicative competence?. *RELC Journal*, 42(1), 856-864.



- Nourdad, N. (2022). Philosophical Approaches to Educational Assessment. *Journal of Philosophical Investigations*, 16(40), 68-85.
- Ockey, G. J., Chukharev-Hudilainen, E., & Hirsch, R. R. (2023). Spoken dialogue systems and their potential for aiding in the assessment of interactional competence. *Language Assessment Quarterly*, 20(5), 377-398.
- Oyinloye, O. M., & Imenda, S. N. (2019). The impact of assessment for learning on learner performance in life Science. *EURASIA Journal of Mathematics, Science and Technology Education*, 15(11).
- Riazi, A. M., & Mosalanejad, N. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's taxonomy. *TESL-EJ*, 13(4), 1-16.
- Skinner, B. F. (1974). *About behaviorism*. New York: Vintage Books.
- Smith, K. (1999). Language testing: Alternative methods. In: Spolsky, B. (ed.), *Concise Encyclopedia of Educational Linguistics*. Amsterdam: Elsevier. Pp. 703-706.
- Stavredes, T. (2011). *Effective online teaching: Foundations and strategies for student success*. John Wiley & Sons.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Weir, C. J. (1990). *Communicative Language Testing*. London: Prentice Hall International (UK).
- Zohrabi, M., & Jafari, H. (2020). The role of think-pair-share interactional activity on improving Iranian EFL learners' willingness-to-communicate. *Teaching English Language*, 14(1), 153-182.
- Zohrabi, M., & Tahmasebi, Z. (2020). A study of the effect of dictogloss as a medium of form-focused instruction on vocabulary versus grammar development of Iranian EFL learners. *Applied Research on English Language*, 9(2), 29-50.
- Zohrabi, M., & Bimesl, L. (2022). Exploring EFL teachers' perceptions of strategies for promoting learners' willingness-to-communicate in online classes. *Applied Research on English Language*, 11(1), 89-110.
- Xiang, R., & Yenika-Agbaw, V. (2021). EFL textbooks, culture and power: A critical content analysis of EFL textbooks for ethnic Mongols in China. *Journal of Multilingual and*

*Multicultural Development*, 42(4), 327–341. <https://doi.org/10.1080/01434632.2019.1692024>.



پرو، شگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی