



APPROACHES TO ASSESSMENT AND TESTING IN FOREIGN LANGUAGE TEACHING AT A NON-LANGUAGE UNIVERSITY

MATIBAEVA Razia Baltabaevna, International Islamic Academy of Uzbekistan,
RYSBEKOVA Lazzat Kasymbekovna International Kazakh-Turkish University named after H.A. Yasawi

Article history:		Abstract:
Received:	24 th September 2023	This article reviews different types of assessment and testing methods with regard to the decisions they help to make and the objectives of foreign language teaching at a non-language university. The article also deals with the practical application of these methods and examines criteria for developing reliable and valid tests. Much attention is given to formative and summative assessment as the most common assessment tools in educational contexts and the importance of their relevance to the second language learning objectives and course content. The article looks at selected responses, constructed responses, and personal responses as generally accepted testing formats and their practical classroom applications.
Accepted:	20 th October 2023	
Published:	28 th November 2023	
Keywords: assessment, testing, validity, proficiency testing, formative testing, summative testing, washback, communicative competence, target language, achievement tests, multiple choice tests, matching test		

Teaching foreign languages at a non-language university involves using different types of assessment and testing methods, which largely depend on the teaching objectives and the students' stage of learning. There are various approaches to assessment, with a general agreement among most researchers in this field that it is beneficial for monitoring learners' progress and achievement in reaching the learning objectives.

Understanding testing and its implications for learning is important to both those involved in test design and those who use testing in educational contexts.

Mihai [13:24] views assessment as a "combination of all formal and informal judgments that occur inside and outside a classroom.»

In an educational setting, given the institutional constraints and varying levels of learners' proficiency, achievement tests are widely accepted as being based on the material covered and directly relating to the curriculum. They are typically used for formative and summative assessments.

At a non-language university, these types can fit into the curriculum at different stages of learning and teaching.

Proficiency testing involves assessing the overall language ability and level of competence across the four skills (listening, speaking, reading, and writing) and is not related to any curriculum. However, it can be used when decisions have to be made if a learner's level of proficiency is sufficient for getting instruction in the target language, be it at a university in their country or abroad.

Placement testing at a non-language university allows for the measurement of learners' abilities with the aim of streaming them into groups of appropriate level of competence. The main purpose here, as Brown indicates, is to allocate them to groups that are "neither too easy nor too difficult, but appropriately challenging.»

Diagnostic tests, as the term implies, are used to diagnose specific areas of language. They identify learners' weaknesses, which should be addressed in further instruction. As Carr [8:7] remarks, they can also be used after placement tests to confirm learners' streaming to appropriate levels.

Achievement tests are designed to show the learners' progress within a certain language program and are based on the material covered. They are normally administered at the end of the course of study and show how the course objectives have been met and measure and grade the performance of individual learners.

Aptitude tests help predict a learner's success before they are exposed to the second language.

Given the institutional constraints of a non-language university, placement and achievement tests are more common as being relevant to the teaching and learning situation. Achievement tests are typically used for formative and summative assessment.

In a non-language university context, formative assessment occurs regularly throughout the semester. It can take different shapes: using grammar tests and vocabulary dictations, assessing the four language skills by a classroom observation, assigning specific communicative tasks to certain students or groups of students within the same class, etc. All these allow the teacher to assess the learners' progress through a definite period of time and work on eliminating their possible language gaps. These assessments are normally not graded.

Conversely, summative assessment measures the success of a language program in the academic area and typically occurs at the end of a module, course strand, or semester. It focuses on the learning outcomes, on how the learners did on a language program, measuring the level of acquired competences and skills. The learners' performance in summative assessment is graded and might inform further educational decisions.

Summative assessment typically includes tasks that enable learners' achievement across the language systems and the four language skills as defined by CEFR [9] and are based on the material covered. When language systems are checked, testing is appropriate, as learners' performance on the test can be promptly measured and communicated to students. The four skills can also be assessed through testing.

Summative assessment produces more objective results when it is administered on an integrated basis and includes tasks that cover all language systems and communication skills. The results inform the teacher of the learners' problem areas that have to be addressed and allow them to align the program to better meet their needs. By evaluating the learners' performance on the test, the teacher can make recommendations to individual learners on how to overcome their language gaps. On the other hand, the consistently high test scores of certain learners allow the teacher to assign them more complex and challenging tasks, thus making the learning process more flexible and individual.

In norm-referenced testing, results of an individual test-taker are compared to the other test-takers' performance. The major limitation here is that the examiners get information on how well test-takers performed with regard to other test-takers, but not, as Carr [8:10] observes, "in an absolute way." It means that it was not their overall ability and competence in language skills that were measured, but their performance with regard to other test-takers.

Conversely, criterion-referenced tests measure test-takers knowledge and skills "in an absolute way" in that they are related to certain educational standards, not to the level of other test-takers. Such assessment establishes the relationship between learners' performance and the body of knowledge to be mastered [13:26].

Another dividing line is between objective and subjective testing.

As the names imply, objective tests are those that can be scored objectively as most selected-response questions have a single correct answer, while subjective tests "involve human judgment" mostly in assessing writing and speaking [8:12]. The former is becoming increasingly popular in computerized assessment, while the latter deals with extended-response questions such as essays, reports, presentations, etc. Subjective tests can generate more than just one correct answer and more ways of formulating it. The distinction between objective and subjective testing is getting somewhat blurred as recent developments in testing methodology have contributed to the growing objectivity of 'subjective' testing. Improving rating scales and assessment criteria to provide objective scoring and grading is one of the main concerns of educational authorities and examination boards.

The limitations of discrete-point and integrated testing are that they only measure test-takers' competence rather than their performance, which caused the need for communicative language testing [14:17]. By the mid-1980s, the methodology of language testing had begun to focus on designing communicative language tests. This means that the need for communicative language testing has been perceived, and much research in this field has been done since then.

The purpose of communicative language testing is to provide the teacher with information about the learners' ability to interact in the target language in specific contexts. Consequently, communicative language tests are designed to measure learners' ability to communicate in real-life situations. They test the four language skills of listening, speaking, reading, and writing and are developed on the basis of communicative competence. Canale and Swain [7:4] break communicative competence into the following four areas: linguistic competence (knowledge of linguistic forms), socio-linguistic competence (the ability to use language appropriately in contexts), discourse competence (coherence and cohesion), and strategic competence (knowledge of verbal and non-verbal communicative strategies).

Apart from testing the above-mentioned communicative competence, this testing approach also allows for measuring learners' ability to use the target language in authentic situations. Under this approach, success in learning or teaching a second language is achieved only when a learner can communicate in the target language by being exposed to authentic listening, speaking, reading, and writing. Therefore, test designers have to prepare tasks that are close to real-life situations outside the classroom.

Brown [5:43] identifies five requirements to be followed when constructing a reliable and valid communicative test: meaningful communication, authentic situations, unpredictable language input, creative language output, and integrated language skills.

Communication is meaningful to students when it draws on their personal experience, thus contributing to their natural language use. Making use of authentic situations can increase the likelihood that meaningful communication will occur.

Authentic situations offer students the opportunity to encounter and use the target language receptively and productively in the real world and demonstrate their language ability and competence.

In communicative assessment, unpredictable language input happens as it is normally not possible to predict what speakers will say. Creative language output occurs in authentic situations when language output is largely dependent on a speaker's input.

Integrated language skills, i.e., a communicative test, will elicit the learners' use of language skills integratively, as is the case in real-life communication [4:21].

Weir [13:II] stresses that "to measure language proficiency... account must now be taken of: where, when, how, with whom, and why language is to be used, and on what topics, and with what effect." All these create challenges for test developers.

Brown [5:11] views performance-based assessment as an effective tool for testing learners in the process of performing "actual or simulated real-world tasks." Learners get exposed to authentic materials, such as newspaper or magazine articles, blogs, videos, films, radio, lectures, etc., and do problem-solving tasks on their own. These can be simulations, role plays, case studies, discussions, etc. A teacher assesses students by observing how they negotiate meaning and interact within groups. Learners are free to use whatever language they have at the task stage. This method of assessment is, however, time-consuming, both for learners, as they need quite a bit of time to prepare, and for teachers, who have to design very specific tasks and guide students at the various stages of their performance. In the context of foreign language teaching at a non-linguistic university, performance-based tasks are primarily used in informal classroom assessment.

Fundamental principles for designing second language assessments include validity, reliability, practicality, and washback.

A test as a measuring tool is said to be valid when it "tests what it is supposed to test" [1:170] in that it measures a learner's ability and competence in a particular area. Weir [14:12] considers a test to be valid if the test scores adequately reflect a learner's ability and communicative competence. Most researchers in this field view validity as the integration of content, criteria, and constructs.

Content validity characterizes a test from the point of view of the content area of the test. A test is valid when [11:22] "its content constitutes a representative sample of the language skills, structures, etc." This aspect of validity ensures the accuracy of testing in terms of its relevance to test specifications and test content.

Criterion-referenced validity is correlated to a certain outside criteria that determines a test's validity. Construct validity is defined by a model of measured quality.

Practicality refers to practical issues at all stages of test design, administration, and assessment [6:19]. A test can be considered practical when it is not expensive and time-consuming and is relatively easy to administer and assess.

Language testing researchers identify three main testing techniques: selected responses, constructed responses, and personal responses.

Selected response tests ask test-takers to choose the correct answer from a certain number of options and include multiple-choice, matching tasks, and true-false statements. As learners do not produce any language themselves, this type of testing is generally accepted for measuring receptive skills such as listening and reading. They can also provide good feedback on learners' abilities in grammar and vocabulary. Practicality is the most obvious benefit of this testing technique, as such tests can be easily administered and scored.

True-false statements ask students to respond to the test questions. Test accuracy can be a factor, as there is a 50% chance of correct guessing. Most language testing researchers believe that this problem can be minimized, provided that true-false questions are carefully constructed and a large number of true-false statements are incorporated into the test. True/false activities are mostly intended to check the learners' reading and listening competences.

Constructed response tests, conversely, are good tools to test learners' productive skills, such as speaking and writing. These can be short answers, interviews, or performance assessments. Such assessments are also appropriate for observing the interaction of productive and receptive skills such as listening and speaking in an interview.

The short-answer format is an alternative to multiple-choice questions, as the learners have to give their own responses. If test questions are properly constructed, the learner will, in high probability, produce short, well-structured answers. A variety of questions can be included in the test, providing more opportunities to assess the learner's performance. However, this testing format involves writing answers and, therefore, may generate vocabulary, grammar, and spelling mistakes.

The potential of testing as a controlling and measuring tool, given its objective character and its positive effect on language teaching and learning, can only be realized when language teachers and test makers gain full awareness of the principles of designing a valid test. Testing the four language skills in different situations, as Galaczi [10] points out, is crucial to increasing test accuracy.

Each method of assessment and testing has its own benefits and drawbacks. No one method can be considered ultimate, and being able to objectively test all language systems and skills. The task of language teachers and test developers is to design testing and measuring materials that are based on the educational content, correspond both to the course objectives and methods of instruction, and, therefore, are more likely to yield reliable results.

REFERENCES:

1. Alderson, C., Clapham, C., and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press
2. Bachman, L. (1990), *Uses of Language Tests* (pp. 58–60) in *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press
3. Baranovskaya, T., & Shaforostova, V. (2017). *Assessment and evaluation techniques* *Journal of Language and Education*, 3(2), 30-38. <https://doi.org/10.17323/2411-7390-2017-3-2-30-38>
4. Brown, D. (2014) *Principles of Language Learning and Teaching*, 6th edition. White Plains: Pearson Education, Inc.

5. Brown, D.; Heekyeong, L. (2015) Teaching by Principles: An Interactive Approach to Language Pedagogy, 4th edition. White Plains: Pearson Education, Inc.
6. Brown, D. (2018). Language Assessment: Principles and Classroom Practices, 3rd edition. White Plains: Pearson Education, Inc.
7. Canale, M., and Swain, M. (1980). Theoretical Bases of Communicative Approach to Second Language Teaching and Testing. *Applied Linguistics*, 1, 1, 1-47.
8. Carr, N. (2011). Designing and Analyzing Language Tests. Oxford: Oxford University Press.
9. Council of Europe: Common European Framework of Reference for Languages <https://www.coe.int/en/web/common-european-framework-reference-languages> retrieved 26/05/2019
10. Galaczi, E. Benefits of Testing the Four Skills (Reading, Listening, Writing, and Speaking), 27/06/2018, <https://www.cambridgeenglish.org/blog/benefits-of-testing-the-four-skills/>, retrieved May 26, 2019.