Language Assessment in Instructional Contexts: Design Considerations and Practical Insights

[1] Roya Zeynalova

Abstract. This article provides a comprehensive overview of various types of language assessments, emphasizing their purposes, design considerations, and practical applications in classroom settings. It first examines less commonly designed tests language aptitude and proficiency assessments highlighting their objectives, tasks, and limitations, particularly in predicting communicative competence. The article then focuses on tests more frequently developed by educators: placement, diagnostic, and achievement tests. Placement tests are explored as instruments for assigning learners to appropriate course levels, with an emphasis on authenticity, diagnostic potential, and instructional alignment. Diagnostic tests are presented as tools for identifying specific learner difficulties, illustrated through detailed phonological and writing assessments. Achievement tests are discussed in relation to curriculum objectives, formative and summative functions, and practical guidelines for their construction. The article also outlines key steps in classroom test design, stressing the importance of clearly defined, performance-based objectives. Overall, the paper underscores the value of informed, purposeful test design in promoting accurate evaluation, learner feedback, and instructional improvement.

Key Words: *Language Assessment, Proficiency Tests, Placement Tests, Diagnostic Tests, Achievement Tests, Test Design*

Introduction: The initial step in constructing an assessment for learners involves identifying its primary purpose. Clarifying the aim of the test not only facilitates the selection of an appropriate test type but also ensures that the assessment remains aligned with clearly defined objectives. At this point, attention will be directed first toward two categories of tests that classroom teachers are less likely to design language aptitude tests and language proficiency tests and subsequently toward three types that educators are more commonly required to develop: placement tests, diagnostic tests, and achievement tests.

Language aptitude tests

One type of test- though relatively uncommon, aim to predict an individual's potential success before actual exposure to a second language. Such tests are constructed to assess a learner's capacity or overall ability to acquire a foreign language, as well as the likelihood of achieving proficiency in that process.

---
[1] Zeynalova, R. Teacher, Nakhchivan State University, Azerbaijan. Email: roya.q90@gmail.com. ORCID: https://orcid.org/0009-0004-8712-7975

In principle, language aptitude tests are intended to be applicable to the classroom learning of any language.

In the United States, two standardized aptitude assessments have been widely employed: the Modern Language Aptitude Test (MLAT) developed by Carroll and Sapon (1958) and the Pimsleur Language Aptitude Battery (PLAB) introduced by Pimsleur (1966). Both instruments are administered in English and require test-takers to complete a series of language-related tasks. For instance, the MLAT is composed of five distinct task types designed to evaluate various aspects of language learning potential.

The Modern Language Aptitude Test (MLAT) incorporates five distinct tasks, each targeting a specific dimension of language learning potential. The first, *number learning*, requires test-takers to acquire a series of numbers presented aurally and subsequently identify various combinations of those numbers. The second, *phonetic script*, assesses the ability to establish associations between speech sounds and their corresponding phonetic symbols. The third component, *spelling clues*, presents words written in a quasi-phonetic form, and examinees must choose from a list the word whose meaning most closely corresponds to the disguised form. The fourth task, *words in sentences*, evaluates grammatical awareness by asking test-takers to match the function of a given key word in one sentence with a word serving the same function in another sentence. Finally, the *paired associates* task requires rapid learning of vocabulary items from a foreign language and the memorization of their English equivalents. (Brown, J. D. 2004).

Research indicates that both the MLAT and PLAB demonstrate notable correlations with students' eventual performance in formal language courses (Carroll, 1981). However, such correlations are largely contingent upon instructional contexts where success is defined through processes such as mimicry, memorization, and problem-solving. There is insufficient empirical evidence to confirm that these tasks serve as reliable predictors of communicative competence, particularly in the context of untutored language acquisition. Due to these limitations, standardized aptitude tests are now rarely employed. Contemporary approaches to assessing language aptitude tend instead to focus on providing learners with insights into their preferred learning styles, as well as their individual strengths and weaknesses, often accompanied by strategies to enhance the former and address the latter. In fact, any test purporting to predict success in second language acquisition is inherently problematic, since research increasingly suggests that with adequate self-awareness, strategic engagement, and appropriate strategy-based instruction, nearly all learners are capable of achieving success. Thus, categorizing learners in advance prior to any actual learning experience risks making unfounded assumptions about their potential for either success or failure.

Proficiency Tests

When the objective of an assessment is to evaluate an individual's overall competence in a language, it is conventionally referred to as a *proficiency test*. Unlike assessments that target a particular course, curriculum, or single linguistic skill, proficiency tests aim to measure comprehensive language ability. Traditionally, such tests have relied on standardized multiple-choice items assessing grammar, vocabulary, reading comprehension, and listening comprehension. In some cases, writing samples are

included, and more recent formats also incorporate tasks involving oral production. Brown, J. D. (2004). Despite longstanding concerns regarding content validity, decades of construct validation research have significantly advanced the development of communicative proficiency assessments.

Proficiency tests are generally summative in nature and norm-referenced, producing a single composite score. This outcome is deemed adequate for their primary gate-keeping function—determining whether a candidate is permitted to advance to the next stage of an academic or professional pathway. Because results are norm-referenced, with equated scores and percentile rankings holding primary importance, such tests are not typically designed to deliver diagnostic feedback. A prominent example of a standardized proficiency test is the *Test of English as a Foreign Language* (TOEFL), developed by the Educational Testing Service. The TOEFL is widely used by over a thousand higher education institutions in the United States as an indicator of an applicant's readiness to engage in academic study within an English-speaking environment. The *Test of English as a Foreign Language* (TOEFL) is composed of several components, including listening comprehension, grammatical accuracy (structure), reading comprehension, and written expression. In addition, the computer-based version of the TOEFL, announced for implementation in 2005, introduced an oral production component. (Educational Testing Service. (n.d.).*TOEFL Test.*) With the exception of the writing section, the TOEFL—similar to many other large-scale proficiency assessments—can be scored by machine, thereby ensuring efficiency, rapid results, and cost-effectiveness, which reflects the principle of practicality. Current research is also exploring the potential of speech recognition technologies to determine whether oral production can be reliably evaluated through automated scoring.

A central challenge in proficiency testing lies in clearly defining the constructs that constitute language ability. Test tasks must represent authentic and contextually relevant samples of English language use. The process of designing such tasks and validating them through empirical research is both labor-intensive and financially demanding. Consequently, it is generally impractical for individual language teachers to attempt the creation of comprehensive proficiency assessments independently. A more feasible approach is to utilize one of the commercially developed and widely available proficiency tests.

Placement Tests

In some cases, proficiency tests may also serve the function of placement tests, which are designed to assign learners to an appropriate level within a language curriculum. Placement tests generally include a representative sample of the materials taught across different courses in the program, so that students' performance indicates the level at which the instructional content will be sufficiently challenging without being either too simple or excessively difficult.An example is the *English as a Second Language Placement Test* (ESLPT) at San Francisco State University, which is structured in three sections. In the first section, students are required to read a short passage and produce a summary essay. The second section involves writing a composition in response to an assigned article. The final section consists of multiple-choice items, in which students read an essay and identify grammatical errors. The maximum duration of the test is three hours. (Fulcher, G.2010).

The rationale behind this tripartite design is primarily supported by content validation. Since most ESL courses at San Francisco State combine reading and writing skills with particular emphasis on writing the test's format mirrors these instructional objectives. The first section simultaneously assesses reading comprehension and writing ability through summary writing. The second section evaluates the ability to express and support opinions, a central requirement of academic writing courses. The third section focuses on error detection, which parallels the proofreading process and reflects an essential academic skill in drafting and revising written work.

Instructors and administrators within the ESL program at San Francisco State University generally view the ESLPT as a more authentic assessment tool compared to its earlier multiple-choice, discrete-point grammar and vocabulary version, largely due to its ability to discriminate effectively among students. Nonetheless, the test poses certain challenges in terms of practicality, as the first two sections require evaluation by human raters a process that is more costly and time-consuming than the automated scoring of multiple-choice responses in Part III. (Brown, J. D. (2004). Concerns regarding reliability also exist; however, these are addressed through careful training of all evaluators. Although the test sacrifices some degree of practicality and reliability, it compensates by providing rich diagnostic information. For example, error analysis from the multiple-choice section reveals students' grammatical and rhetorical weaknesses, while the essay tasks serve as a valuable preview of learners' writing abilities for their future instructors.

Placement tests, more broadly, exist in diverse formats. They may assess both comprehension and production, incorporate written or oral performance, and employ open-ended responses, restricted responses, selection tasks (such as multiple-choice), or gap-filling activities, depending on the specific goals and requirements of a given program. Some institutions opt for standardized proficiency tests as placement instruments due to their practicality namely, lower costs, faster scoring, and efficient reporting. Others prioritize the more detailed performance data yielded by open-ended written or oral production tasks. Ultimately, the central purpose of any placement test is to ensure accurate placement of learners within appropriate courses or levels. Additional advantages to consider include face validity, the provision of diagnostic insights, and the authenticity of the assessment tasks. (Fulcher, G.2010).

During a recent one-month summer program in English conversation and writing at San Francisco State University, 30 students were required to be assigned to one of two course sections. The placement test employed for this purpose consisted of a five-minute oral interview and an essay-writing task, with the primary goal of dividing the students into two evenly balanced sections based on performance. While this objective could have been accomplished through a simple, grid-scored multiple-choice grammar and vocabulary test, the inclusion of the oral interview and writing component offered additional advantages. Specifically, these tasks enhanced the test's face validity, provided a more personalized approach suitable for a small program, and generated valuable diagnostic information about the students, most of whom were previously unknown to the instructors.

Diagnostic Tests

A diagnostic test is designed to identify specific aspects of language proficiency that require attention. For instance, a pronunciation diagnostic may evaluate phonological features of English that learners

find particularly challenging, thereby informing curriculum development. Typically, such tests provide a checklist of features for the administrator often the instructor to systematically identify areas of difficulty. Similarly, a writing diagnostic collects student writing samples to help teachers pinpoint rhetorical and linguistic elements that warrant focused instruction. (Clifford Prator (1972).

As previously noted, diagnostic and placement tests can sometimes overlap. For example, the San Francisco State ESLPT serves both placement and diagnostic functions. Placement tests that deliver information beyond simple course-level assignment may concurrently fulfill diagnostic purposes. A distinction also exists between diagnostic tests and general achievement tests: achievement tests measure the extent to which students have mastered previously taught content, whereas diagnostic tests aim to reveal areas requiring future instruction. Consequently, diagnostic assessments typically provide more granular, subcategorized information about learners. In curricula with a form-focused component, for example, a diagnostic test may yield insights into a learner's mastery of verb tenses, modal auxiliaries, definite articles, relative clauses, and similar grammatical structures.

An illustrative example of a diagnostic test for oral production is Clifford Prator's (1972) assessment developed alongside an English pronunciation manual. Test-takers read a 150-word passage while being tape-recorded. The administrator then uses a detailed inventory of phonological items to analyze the learner's performance. After repeated listenings, the administrator completes a checklist of errors organized into five main categories, each comprising multiple subcategories.

The primary categories in Prator's (1972) diagnostic test include:

 (1) stress and rhythm,

(2) intonation,

(3) vowels,

 (4) consonants, and

(5) other relevant phonological factors.

For instance, the *stress and rhythm* category is further divided into subcategories such as:

(a) stress placed on an incorrect syllable in multi-syllabic words,

(b) improper sentence stress,

(c) inaccurate segmentation of sentences into thought groups, and

(d) inadequate transitions between words or syllables.

This detailed information enables instructors to determine which aspects of English phonology require focused instruction. Additionally, it provides learners with valuable feedback, raising their awareness of specific errors and promoting the development of appropriate compensatory strategies.

Achievement Tests

An achievement test is directly linked to the content of classroom lessons, units, or an entire curriculum. Such tests are designed to assess material that has been explicitly covered within a specified instructional timeframe and are typically administered after students have engaged with the relevant objectives. While achievement tests can provide diagnostic insights into areas requiring further study, their primary purpose is to determine whether students have met course objectives and acquired the targeted knowledge and skills by the end of the instructional period.Achievement tests are often summative, administered at the conclusion of a unit or term, but they can also serve an important formative function. A well-constructed achievement test can provide washback, offering feedback on learners' performance in specific subsets of the unit or course, thereby contributing to ongoing instructional improvement. The design of an achievement test should be guided by several factors: (Weir, C. J. (2005).

 (1) the objectives of the lesson, unit, or course being evaluated;

(2) the relative importance or weighting of each objective;

 (3) the types of tasks employed during classroom instruction;

(4) practical considerations such as testing duration and scoring turnaround; and

(5) the extent to which the test format facilitates formative feedback and instructional washback.

Some practical steps to test design

What is the primary purpose of a test? While it is improbable that one would be tasked with designing a language aptitude or proficiency test, understanding their characteristics is essential for accurately interpreting their results. In contrast, teachers are likely to have numerous opportunities to develop placement, diagnostic, and particularly achievement tests.It is a common misconception that every test must be an exceptionally innovative instrument capable of earning acclaim from colleagues or admiration from students. In reality, designing novel assessment formats requires substantial effort and iterative refinement through trial and error. (Bachman, L. F., & Palmer, A. S. (1996). Moreover, traditional testing methods, when applied creatively, can effectively align with the principles of an interactive and communicative language curriculum. For novice teachers, the most practical approach is to adhere initially to established and widely accepted testing practices. With experience, educators can gradually experiment with more innovative assessment strategies. Within this context, it is valuable to consider practical steps for constructing effective classroom tests.

Assessing Clear and Specific Objectives

In addition to understanding the overall purpose of a test, it is essential to identify precisely what is intended to be assessed. Frequently, teachers may administer tests arbitrarily perhaps because it is Friday of the third week of the course creating items hastily from a quick review of the material covered. Such an approach is pedagogically unsound. Instead, assessment should begin with a careful examination of all content and skills that students are expected to master, based on the unit objectives. (Hughes, A. (2003).

Curricula should define assessable objectives in terms of observable student performance. An objective that merely states "Students will learn tag questions" or simply labels a grammatical focus as "Tag questions" is insufficiently specific for testing purposes. Such statements fail to clarify whether students should comprehend these structures in spoken or written form, produce them orally or in writing, or use them in particular contexts such as conversation, essays, or academic lectures. Therefore, the first step in test design is to determine objectives that are appropriately specified and measurable. (Brown, J. D. 2004).

For example, consider a low-intermediate integrated-skills class emphasizing social conversation and incorporating some reading and writing activities. The objectives for this unit either already established or reformulated by the instructor should be articulated in terms of the performance expected from students and the targeted linguistic domains. Each objective should thus provide clear guidance regarding what students should be able to do and in which contexts, ensuring that the assessment accurately reflects the intended learning outcomes.

Form-focused objectives (listening and speaking)

Students are expected to:

1. Identify and accurately produce tag questions in simple social conversations, demonstrating correct grammatical structure and appropriate final intonation patterns.

2. Recognize and correctly formulate wh-information questions, applying the appropriate final intonation in spoken interactions.

Communication Skills (Speaking)

Students are expected to:

3. Describe completed actions and events accurately within social conversations.

4. Request confirmation or clarification appropriately during interactions.

5. Express opinions about events in social conversations with clarity.

6. Produce spoken language that demonstrates contextually appropriate intonation, stress, and rhythm.

Reading Skills (Simple Essay or Story)

Students are expected to:

7. Identify and correctly recognize irregular past tense forms of selected verbs within a story or short essay.

Writing Skills (Simple Essay or Story)

Students are expected to:

8. Compose a one-paragraph narrative describing a simple past event.

9. Appropriately employ the conjunctions so and because when expressing opinions in writing.

Conclusion

Effective language assessment requires a clear understanding of the purpose, scope, and objectives of each test. While language aptitude and proficiency tests provide valuable insights into learners' potential and overall competence, their practical application in classroom settings is limited. Conversely, placement, diagnostic, and achievement tests are directly relevant to instructional contexts, offering critical information for student placement, identification of learning needs, and evaluation of curriculum effectiveness. The design of these assessments should be guided by clearly defined, performance-based objectives, ensuring alignment with targeted skills and linguistic domains. Additionally, a balanced approach that combines traditional testing methods with innovative, communicative tasks can enhance both the authenticity and instructional value of assessments. Ultimately, well-constructed language tests not only measure learner performance accurately but also support ongoing learning by providing diagnostic feedback, promoting learner awareness, and facilitating strategic instructional planning. Careful attention to test design, clarity of objectives, and alignment with curriculum goals is essential for fostering meaningful, reliable, and effective language assessment practices.

References

Brown, J. D. (2004). *Language Assessment: Principles and Classroom Practices*. Pearson Education.

Carroll, J. B. (1981). *Twenty-five years of research on foreign language aptitude*. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.

Carroll, J. B., & Sapon, S. M. (1958). *Modern Language Aptitude Test*. Language Learning and Testing Foundation.

Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery*. Language Learning and Testing Foundation.

Clifford Prator (1972). *A Diagnostic Test of English Pronunciation*. Prentice-Hall.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press.

Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.

Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press.

Educational Testing Service. (n.d.). *TOEFL Test*. Retrieved from https://www.ets.org/toefl