Interdisciplinary Research in Education
Volume 9, Issue 2, 2024: 26-37
DOI: https://doi.org/10.3126/ire.v9i2.75020

From Traditional to Modern: Evolving Practices in Language Testing and Assessment

Gopal Prasad Pandey
Associate Professor, English Education, Central Department of Education, Tribhuvan University
ORCiD: https://orcid.org/0000-0003-1671-0501
Email: gpandeytu@gmail.com

Keywords

Alternative assessment, authentic assessment, dynamic assessment, task-based language assessment, critical language testing

Abstract

This paper looks into the shifts from traditional to innovative and inclusive language testing methods, focusing on developments such as alternative assessment, AI integration, and computer-adaptive testing. This study employs a desk-based approach to the literature which synthesizes relevant theories in order to highlight the increased focus on fairness, inclusivity and authentic assessment. The key findings highlight that dynamic and performance-based assessments address the different learning needs of the learners; promote instrumentality towards language use and incorporate socio-political and ethical dimensions in test construction. The paper also underlines the transformative power of technology to improve access and efficiency considering the potential challenges and inequalities associated with its use. By integrating traditional and modern practices, this study contributes to both the theoretical discourse and practical advancements in language assessment that uphold equitable, ethical and effective testing methods consistent with twenty-first century educational standards and learner diversity.

Introduction

Language testing has evolved significantly in recent years, reflecting shifts in educational paradigms, technological advancements and the growing emphasis on contextual and ethical considerations. It is essential for applied linguists and language teachers to have a clear understanding of language testing. Davies (1982) states, "Language testing has come of age and is now regarded as providing a methodology that is of value throughout applied linguistics and perhaps in core linguistics too" (p. 141). Fulcher and Davidson (2012) argue that "a failure to design or select appropriate instruments threatens the validity of the research throughout applied linguistics and second language acquisition" (p.1). A notable trend is the increasing focus on alternative assessments, such as portfolios, self-assessments and peer assessments which aim to provide a more holistic view of a learner's language abilities (Bachman & Palmer, 2010). The incorporation of technology, including artificial intelligence

(AI) and natural language processing (NLP), in language assessments has created innovative methods for evaluating and conducting online examinations. These assessments provide immediate feedback (Maeda 2023). This technology influences our evaluation of language proficiency.

Test-takers' views on AI incorporation into language testing highlight another important aspect that Zhang et al. (2023) sought to bring to the forefront-the need for fairness and ethical considerations in AI-mediated assessment. Caines et al. (2023) also examined the opportunities and challenges posed by LLM adoption in language teaching and assessment and their associated implications for educational technology. Moreover, there is a growing trend toward assessment for learning (AfL) which, in contrast to the previous concept of assessment of learning, describes an approach where assessment is formative rather than summative and serves the purpose of enhancing learning instead of just being treated as a post-performance evaluation Black & Wiliam, 1998.

One important consideration is that tests need to be adapted to cater for different needs, in particular, test accommodations (for people with disabilities) are necessary to ensure inclusivity and fairness while testing (Kunnan, 2004). Furthermore, critical perspectives on language testing interrogate social and ethical concerns, and speak to the rights and responsibilities of both test takers and test designers. Big data analytics are transforming the field of scale of tests performance analysis (test reliability and validity). Such developments occur against a backdrop of broader trends toward ensuring that language testing is not only valid and reliable, but also to ensure that they are attuned to the values of fairness, inclusivity and meaningful learning, while using new technologies in appropriate ways.

Recent discussions about language testing center on growing demands for fair and inclusive ways to assess people, considering the wide range of language backgrounds test-takers have. While traditional language tests have spread, people now question if they're valid and reliable for addressing the language diversity and cultural specifics of learners, as the world becomes more multilingual. Also, using tech in language assessment has brought more attention to digital equality and fairness as well as how automatic systems might repeat biases. This has led to calls for testing methods that are more open and responsible, which can adjust to the changing needs of a global society.

The paper aims to discuss the future of language testing with a focus on what new practices are emerging and how traditional language testing methods are being transformed to stay relevant in an ever-changing world. This also analyzes key advancements such as alternative assessments, integration of technology like artificial intelligence and assessment for learning, which promote a holistic understanding of language proficiency. Furthermore, the study examines the socio-political and ethical factors that are involved in testing, such as justice, inclusion, and the consequences of large-scale testing.

Review of Theoretical Approaches to Language Testing and Assessment

Communicative Language Testing Theory (CLTT), developed as a reaction to traditional testing methods, emphasizes the assessment of learners' ability to use language in authentic communicative contexts. Based on Canale and Swain's (1980) four components of communicative competence-grammatical, discourse, sociolinguistic and strategic

competence, it places emphasis on practical applications. Bachman (2010) further expanded this framework by identifying language competence, strategic competence and psychophysiological mechanisms as its three main components. Constructivist approaches rooted in Vygotskian points of view (1978) emphasis developing responsibility and active meaning-making on the part of students. Formative assessments such as peer reviews and portfolios empower students. Authentic assessment techniques (Wiggins, 1993) help projects that reflect real-life occurrences, thereby linking classroom learning to practical real-world communication. Task-Based Language Assessment (Ellis, 2003) emphasizes real-life tasks for holistic assessment of integrated skills and dynamic assessment (Vygotsky, 1978), which conceptualizes testing and teaching as an inseparable quantity as it occurs within the Zone of Proximal Development (ZPD) to gauge the learner's current performance and potential performance.

Computer-Adaptive Testing (CAT) is an innovative tool with a technological approach that changes the difficulty level of test items according to the test taker's response in real-time. This is an adaptive testing procedure based on Item Response Theory (IRT), which certifies each test-taker is individualized in the assessment experience. Alternative assessment theory focuses on using formative measures (e.g., portfolios) to represent the development of learners more broadly (Brown & Hudson, 1998). Washback and impact theory focus on the consequences that tests have on instruction, learning and the educational system at large (Alderson & Wall, 1993). Washback refers to the influence of language assessment on language learning, teaching, and instruction, which can be either beneficial or detrimental; positive washback promotes effective teaching and learning, while negative washback leads to adverse outcomes such as rote memorization. Hughes (2010) states, "the effect of testing and learning is known as washback, and can be harmful or beneficial" (p.1). Theoretical frameworks for multimodal assessment (Kress, 2010) emphasize various pedagogical modes of communication, representative of language in use today. The socio-cognitive theory of test validation uses a socio-cognitive approach to evaluate language assessments (Weir, 2005). The test usefulness put forth by Bachman and Palmer (1996), states six important test qualities: reliability, construct validity, authenticity, interactivity, impact and practicality. These attributes inform the development of balanced and effective language assessments. Another theory that represents the ethical aspect of language testing, which has been mainly approved by the fairness and social justice theories, will lead to an emphasis on equal test design and implementation. This includes ensuring that test accommodations are provided for test-takers with disabilities, and that there are no cultural biases in the test items.

Methodology

The study adopts a qualitative, desk-based approach to the literature to determine how language testing has changed and where it is going. It encompasses diverse facets of the progression of the field, through synthesizing a body of scholarly works from foundational theories to empirical studies to contemporary advancements. It critically analyses historical approaches (essay-translation, structuralist, integrative and communicative) along with contemporary trends (alternatives to current assessment, computer-adaptive testing and the integration of artificial intelligence (AI). The study is grounded in established theoretical frameworks, including the socio-cognitive

theory of test validation, task-based language assessment, and washback and effect theory, thereby framing the debate. From a critical perspective, it explores the sociopolitical and pragmatic consequences of language testing policies addressing fundamental concerns including justice, inclusion, and ethics. Insights from the fields of applied linguistics, educational psychology, and technology studies provide an interdisciplinary perspective that informs the discussion, especially regarding the advantages and disadvantages of AI and dynamic assessments.

Results and Discussion

This section synthesizes insights derived from desk-based literature analysis on the changing landscape from old school to new school assessment practices. It explores key themes, including the evolution of assessment methods from traditional to modern practices, technology integration, and the increasing focus on fairness, inclusivity, and authenticity. The discussion looks at both the theoretical and practical advancements and their impacts on education advocating for the alignment of language assessments with a diverse, technology-oriented, and equity-centred world.

Evolving Perspectives on Assessment in Education

Heaton (1988) states that there are four basic techniques to language testing: (i) the essay-translation approach, (ii) the structuralist approach, (iii) the integrative approach, and (iv) the communicative approach. Even though these methods are presented in chronological sequence, they should not be seen as being limited to specific points in the history of language testing. Moreover, the approaches are not always entirely distinct from one another. An effective assessment will typically integrate elements from multiple methodologies. Indeed, a test may possess intrinsic limitations due to its reliance on a singular methodology, regardless of how appealing that methodology may seem.

The Essay-Translation Approach. This type of approach is sometimes called the pre-scientific stage in the history of language testing. It does not call for specialized skills or expertise in testing, as the judgment of the teacher as such is considered paramount. Tests typically involve such tasks as essay writing, translation, and grammatical analysis, frequently including comments on the language studied. Tests often reflect a strong literary and cultural emphasis. In public examinations, such as secondary school leaving tests, an aural/oral component may be included at upper intermediate and advanced levels.

The Structuralist Approach. This method stems from the conviction that learning a language primarily requires developing systematic habits. Emphasizing contrastive analysis and the requirement to evaluate students' understanding of various components of the target language, such as phonology, vocabulary, and grammar, it largely borrows from structural linguistics Under this method, testing usually entails the efficient covering of a larger spectrum of language elements by means of isolated words and phrases taken from context. The emphasis is on assessing one ability at a time, hence skills including listening, speaking, reading, and writing are evaluated independently. Although these aspects are occasionally attacked, they are nevertheless important for some kinds of exams, especially those meant to isolate particular skills, including assessing writing without depending on reading comprehension.

The Integrative Approach. This method concentrates on evaluating language in

context, emphasizing the significance of the language and the overall effect of the discourse. Integrative assessments are designed to evaluate a learner's capacity to use multiple skills simultaneously unlike approaches that isolate language skills in order to improve the reliability of the test. They adopt a holistic approach to language competency, based on the idea that learners have an innate language competence or "grammar of expectancy," no matter what their individual learning objectives are. Although integrative testing is based on the idea of "functional language," it does not mean that functional language must be used directly. Cloze testing and dictation are the approaches that are most typically linked with integrative tests.

The Communicative Approach. The communicative method to language testing is sometimes linked to the integrative approach since they both place emphasis on the meaning of utterances rather than their form or structure. On the other hand, the two methods are essentially different. Communicative tests focus on evaluating how language is utilized in real-life situations. As a result, these assessments frequently include assignments that closely resemble circumstances that students may come across outside of the classroom. Success is determined by how effectively language is used, not by how closely it follows formal linguistic rules. This method focuses on language "use," which refers to how individuals use language for different reasons, rather than "usage," which refers to the formal patterns of language as specified by prescriptive grammars and lexicons.

Misconceptions about Language Testing

Misconceptions about language testing often stem from oversimplified or unrealistic beliefs about the nature and purpose of tests. One common misconception is the belief that there is a single "best" test for any given situation, ignoring the fact that testing needs vary based on context, objectives, and learner profiles Bachman and Plamer (1996, p.7). Another misunderstanding lies in the nature of language testing and its development, where the complexities of designing reliable and valid tests are underestimated. Additionally, many hold unreasonable expectations about what language tests can achieve, such as expecting them to provide definitive measures of overall proficiency without considering their limitations. Lastly, placing blind faith in measurement technologies can lead to overreliance on tools and statistical techniques, overlooking the critical judgments and contextual factors that are crucial for meaningful language assessment.

Alternative Assessments

Tests have become an infallible measure in our culture, especially in education. However, research in the 1990s argued against the notion that all people and skills could be measured by traditional tests, leading to the emergence of alternative assessment. However, “research and practice during the 1990s provided compelling arguments against the notion that all people and all skills could be measured by traditional tests” (Brown & Abeywickrama, 2018, p. 16). This shift led to the development of what is now referred to as alternative assessment. This alternative approach consists of additional measures of students, such as portfolios, journals, observations, self-assessments, and peer assessments. Some argue that these alternatives have ethical potential in promoting fairness and balance of power in the classroom.

Alternative assessments in language testing are multiple means of measuring a learner's language skills that go beyond traditional standardized tests. These evaluations focus

on real-world language use by prioritizing realistic, performance-based tasks, such as portfolios, projects, presentations, and peer assessments. The aim is to give a more comprehensive understanding of a learner's language competency and communication skills, with an emphasis on their capacity to use language successfully in different situations. Teachers and educators can enable personalized instruction, accommodate to unique learning styles, and create a deeper engagement with the language learning process by introducing alternative assessments.

Alternative assessments can take many forms, from portfolios and project-based assessments to self-assessment and peer assessment. These assessments emphasize authentic, performance-based tasks that mirror real-life community practices of the language such as portfolios, projects, presentations and peer assessments. It aims to offer a broader perspective on an individual's linguistic capabilities, centering on their capacity to communicate successfully in diverse situations.

Alternative assessments in language testing refer to diverse methods of evaluating a learner's language ability beyond traditional standardized tests. These assessments prioritize authentic, performance-based tasks that reflect real-world language use, such as portfolios, By incorporating alternative assessments, educators can support differentiated instruction, cater to individual learning styles, and promote a deeper engagement with the language learning process.

Alternative assessment refers to a variety of evaluation methods that diverge from traditional standardized tests to assess student learning, performance, and progress. These approaches emphasize the importance of direct, authentic demonstrations of skill and knowledge in real-world or applied contexts. Alternative assessments are often used to complement or replace traditional tests, providing a more holistic picture of student abilities and learning outcomes.

Portfolios. Portfolios are collections of student work that demonstrate learning progress, effort, and achievement over time. They can include a wide range of materials, such as written assignments, projects, art, and reflective essays.

Performance Assessments. Performance assessments require students to perform a task or set of tasks that demonstrate their knowledge, skills, and competencies. These can range from oral presentations and debates to scientific experiments and artistic performances.

Project-Based Assessments. In project-based assessments, students undertake extended projects that require research, planning, and execution over time. These projects often address real-world problems or questions, requiring students to apply interdisciplinary knowledge and skills.

Self and Peer Assessments. These involve students evaluating their own work or that of their peers against a set of criteria or rubrics. Self and peer assessments encourage reflection, critical thinking, and a deeper understanding of learning objectives and standards. Self and peer assessment can enhance learning by promoting metacognitive skills and encouraging students to take responsibility for their learning.

Observational Assessments. Teachers or educators observe students in the process of learning or completing tasks, often using checklists or rubrics to record observations. Observational assessments are particularly useful in early childhood education and in assessing skills that are difficult to measure through traditional tests, such as social interaction and motor skills. Observational assessments in providing immediate, context-specific feedback to support student learning.

Table 1 highlights the differences between traditional test designs and alternatives that are more authentic in eliciting meaningful communication. However, it is difficult to draw a clear line between traditional and alternative assessments.

Table 1

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 sufficient for feedback	Individualized feedback and washback
Norm-referenced scores	Criterion-referenced scores
Focus on discrete answers	Open-ended, creative answers
Summative	Formative
Oriented to product	Oriented to process
Noninteractive performance	Interactive performance
Fosters extrinsic motivation	Fosters intrinsic motivation

(Brown & Abeywickrama, 2018, p. 22

Assessment methods can be categorized into two types: standardized and alternative. Some forms combine both, while others are more subjective and individualistic. The table suggests that assessment traditions should be valued and utilized for their functions. However, alternative assessment methods can be constructively used in classrooms. While more time and institutional budgets are required for subjective evaluations, they provide more useful feedback, intrinsic motivation, and a more complete description of a student's ability.

Performance- Based Assessment

Performance-Based Assessment over the past two decades, a growing number of educators and advocates for educational reform have called for reducing the emphasis on large-scale standardized tests in favor of contextualized and communicative assessments that more effectively support learning in schools. The movement toward what is now referred to as performance-based assessment (Shohamy, 1995) aligns with broader educational reform efforts that strongly oppose relying solely on standardized test scores as measures of student competencies (Lane, 2010). The argument was that standardized tests do not elicit actual performance on the part of test-takers. “The argument was that standardized tests do not elicit actual performance on the part of test-takers” (Brown & Abeywickrama, 2018, p. 18). Performance-based assessment in language generally includes “oral production, written production, open-ended responses, integrated performance (across skill areas), group performance, and other

interactive tasks" (Brown & Abeywickrama, 2018, p. 18). A defining feature a (though not all) performance-based language assessment is the inclusion of interactive tasks, which is why they are also referred to as task-based assessments.

Assessment for Learning

The notion of Assessment for Learning (AFL) is a paradigm shift in education which focuses on the role of assessment to support and improve teaching and learning instead of just measuring it. In contrast to summative assessments (evaluating learning at the conclusion of an instructional period), AFL infuses formative assessments into the learning process to provide real-time feedback for both teachers and students. Continuing assessment gives feedback for instruction and engages students in their learning process. Types of practice in AFL, for example, observations, peer assessment, and self-assessment, provide regular feedback on performance, facilitating the setting of goals and how students may achieve these goals. AFL promotes of cooperation among students in which they have the freedom to incorporate new information from feedback. This is a part of learning process. Creating meaningful tasks and feedback in a timely way is not the only aspect teachers need to deal with effective AFL in practice; students should similarly work actively with feedback to constructively get feedback and track his/her progression. Table 2 provides a detailed overview of the defining characteristics and essential elements of AFL, highlighting its role in fostering meaningful and effective learning experiences.

Table 2

Key Aspects of Assessment for Learning (AFL)

Aspect	Description
Purpose of AFL	Focuses on enhancing learning, not just measuring achievement.
Formative Approach	Uses ongoing feedback integrated into the learning process.
Student Involvement	Encourages active participation, self-assessment, and peer feedback.
Collaborative Learning	Creates a culture where feedback is constructive and errors are seen as learning opportunities.
Teacher's Role	Involves designing assessments, providing feedback, and adapting teaching methods.
Skill Development	Considers test scores as potentially prescriptive but open to discussion.
Integration in Curriculum	Embeds assessment as a continuous part of the teaching-learning cycle.
Comparison to AoL	Contrasts AFL's focus on learning improvement with the evaluative role of summative assessment (AoL).

Test Accommodations

Test accommodations refer to changes in the testing environment, procedures, or materials designed to help individuals with disabilities or other special needs take tests without an unfair disadvantage. The goal of the changes is to reduce barriers that would prevent people from fully participating in and demonstrating their knowledge because of their disabilities-without at the same time changing the fundamental nature and purposes of the test. The purpose of accommodations is to

level the playing field so that what is being measured is the knowledge or skills intended and not the person's impairment.

Focus on Test Usefulness

The focus on test usefulness has become a significant trend in language testing which emphasizes practical applications and impact of tests in real-world contexts. According to Bachman and Palmer (1996), the usefulness of a test is determined by its reliability, validity, authenticity, interactivity, practicality and positive washback. Test usefulness refers to "the purpose of an assessment or the extent to which a test accomplishes its intended criterion or objective" (Brown & Abeywickrama, 2018, p. 61). Recent developments prioritize designing assessments that not only measure language proficiency effectively but also enhance teaching and learning practices. For instance, tests are increasingly evaluated for their impact on stakeholders, ensuring they support educational goals and provide meaningful feedback for improvement. This shift highlights the importance of aligning test purposes with learners' needs and educational outcomes, making assessments a more integral part of language development.

Critical Approach to Language Testing

The critical approach to language testing is a perspective that scrutinizes the broader socio-political implications of language assessments. It goes beyond evaluating the technical aspects of tests, such as reliability and validity, to consider how tests impact societal structures, individual identities, and power distributions. This approach is informed by critical theories in education, which highlight the ways in which language testing can perpetuate inequalities and reinforce dominant cultural norms. Proponents of this viewpoint argue that language tests often serve gatekeeping roles, deciding who has access to educational and professional opportunities. By critically examining the uses and consequences of language tests, this approach seeks to foster more equitable testing practices that are sensitive to the diverse backgrounds and needs of test takers.

Viewing tests in reference to social, educational and political contexts situates the field of testing in the domain of critical testing. In reference to language testing, it is referred to here as critical language testing (Shohamy, 1998, p. 332). Critical language testing presumes that the act of testing is not neutral. Rather, it is both a product and a facilitator of cultural, social, political, educational, and ideological objectives that influence the lives of individual participants, teachers, and students. According to Shohamy (1998), the critical approach to language testing entails the following key features:

Critical language testing views test takers as political subjects situated within a broader political context, emphasizing that language tests function as tools deeply embedded in cultural, educational, and political spheres where ideological and social struggles for dominance take place. It raises questions about the agendas conveyed through tests, their origins, and whose interests they serve. (pp. 332-33)

Challenging traditional psychometric approaches, it adopts interpretive perspectives, urging language testers to critically reflect on the societal visions created or supported by language tests-whether they merely fulfill predefined curricular or proficiency goals or serve broader ideological purposes. It also questions the nature of knowledge underlying tests, asking whether the content represents unquestionable "truth"

or is open to negotiation, challenge, and appropriation. Furthermore, it examines the meaning of test scores, their degree of finality or prescriptiveness, and their openness to interpretation. Rejecting the notion of a test as an isolated tool, critical language testing asserts that language testing is inextricably linked to the educational and social systems in which it operates, making the concept of "just a test" untenable. Critical language testing can be seen as a source of continued and enhanced social dialogue and debate about language testing and its forms and practices, its relation to language teaching and learning, and the roles of the language testers. This incorporates new validity criteria for language testing, such as consequential, systemic, interpretive and ethical validity.

Table 3 outlines the key features associated with critical language testing, highlighting its role in addressing these critical dimensions.

Table 3

Features of Critical Approaches to Language

Features	Description
Political context	Views test takers as political subjects influenced by broader political, cultural, and social systems
Questioning Agendas	Explores whose agendas are embedded in tests and what ideologies they serve.
Challenge to Psychometrics	Moves from traditional psychometric approaches to interpretive perspectives.
Knowledge as Negotiable	Questions whether test content is absolute truth or open to challenge.
Impact on Society	Examines the societal vision and implications created by language tests.
Scores as Open to Interpretation	Considers test scores as potentially prescriptive but open to discussion.
Integration with Social Systems	Recognizes tests as deeply tied to broader social and educational systems.

Use of Technology in Testing

Technological innovation has significantly transformed language learning and teaching, making tools such as smartphones, tablets, and computers integral to modern pedagogy. This evolution is evident in the widespread adoption of computer-assisted language learning (CALL) and mobile-assisted language learning (MALL), which provide learners with interactive, flexible, and personalized learning experiences. Language assessment has also kept pace with these technological advancements, incorporating sophisticated systems like computer-based and computer-adaptive tests (CAT). CATs, for instance, dynamically adjust the difficulty of questions based on a learner's performance, ensuring a tailored testing experience. High-stakes exams like the TOEFL and Pearson Test of English (PTE) now include automated scoring for essays and oral production, which enhances efficiency and standardization. These innovations not only streamline the assessment process but also make it more accessible, as tests can be administered on a global scale with rapid result generation.

Despite the benefits associated with technology-assisted testing, there are a number of limitations that should be

accounted for. Test security and the reliability of assessments offered via informal online quizzes may present some problems. Furthermore, multiple-choice formats often do not provide for a complex evaluation of language. It is important to integrate an opportunity to assess oral or written communication skills that can be challenging without human interaction into the testing design. There is also a chance that technical issues including software malfunctions or unequal access to digital resources would present additional barriers. Nevertheless, the possibility to use technology in testing is committed to language assessment professionals, and it is now their duty to find out how to make it advantageous.

Conclusion

This study aimed to investigate the evolving practices in language testing, emphasizing the shift from traditional methodologies to innovative and inclusive approaches. Adopting a descriptive methodology, it integrated foundational theories, empirical research and contemporary technological advancements. The study highlights that the growing adoption of alternative assessments, such as portfolios and self-assessments and the integration of AI-driven tools emphasize inclusivity, fairness and dynamic feedback mechanisms. The paper concluded with discussions on socio-political implications of language testing and the importance of good practices and increased transparency in the design and use of such tests. These findings highlight the need for testing methods to be aligned with contemporary paradigms of learning to improve validity, reliability and fairness. This finding is important both conceptually and practically in language assessment. Based on the findings, theoretical implications contribute to communicative language testing theory, while pedagogical implications focus on the role of formative assessments in promoting learners' autonomy. However, there are limitations of AI that could foster bias in technology-powered evaluations and accessibility issues, which indicate a need for improvement. By bridging theoretical insights with practical applications, this study paves the way for more inclusive and effective language testing practices.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Bachman, L. F. (2010). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653-675.

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O., Yuan, Z., Elliott, M., Moore, R., Bryant, C.,

Rei, M., Yannakoudakis, H., Mullooly, A., Nicholls, D., & Buttery, P. (2023). *On the application of large language models for language teaching and assessment technology*. *arXiv Preprints.*

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.

Davies, A. (1982). Language testing 2. In A. Kinsella (Ed.), *Surveys 1: Eight state-of-the-art article on key ideas in language teaching* (pp.141-159). Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford University Press.

Fulcher, G., & Davidson, F. (2012). *The Routledge handbook of language testing*. Routledge.

Heaton, J.B. (1988). *Writing English language tests* (2nd ed.). Longman.

Hughes, A. (2010). *Testing for language teachers* (2nd ed.). Cambridge University press.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication.* Routledge.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *Studies in language testing 18: European language testing in a global context* (pp. 27-48). Cambridge University Press.

Lane, S. (2010). Performance assessment: The state of the art. In M. K. Rothman & L. L. Hamilton (Eds.), *Testing and assessment in the 21st century: Current research and practice* (pp. 67-90). Routledge.

Maeda, H. (2023). Field-testing items using artificial intelligence: Natural language processing with transformers. *arXiv Preprints*.

Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics, 15*, 188-211. https://doi.org/10.1017/S026719050000266X

Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation, 24*(4), 331–345. https://doi.org/10.1016/S0191-491X(98)00018-9

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach.* Palgrave Macmillan.

Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* Jossey-Bass.

Zhang, D., Hoang, T., Pan, S., Hu, Y., Xing, Z., Staples, M., Xu, X., Lu, Q., & Quigley, A. (2023). Test-takers have a say: Understanding the implications of the use of AI in language tests. *arXiv preprint arXiv:2307.09885*. https://arxiv.org/abs/2307.09885