

MENTAL ENLIGHTENMENT SCIENTIFIC – METHODOLOGICAL JOURNAL



MENTAL ENLIGHTENMENT SCIENTIFIC – METHODOLOGICAL JOURNAL

<http://mentaljournal-jspu.uz/index.php/mesmj/index>



PECULIARITIES OF READING TEST DEVELOPMENT FOR ESP STUDENTS

Umida Mamatyusupovna Abduvakhobova

University of World Economy and Diplomacy

Umida4080@gmail.com

Tashkent, Uzbekistan

ABOUT ARTICLE

Key words: test development, reading assessment, authenticity, task design, language testing methodology, assessment specification.

Received: 10.12.25

Accepted: 11.12.25

Published: 13.12.25

Abstract: The development of reading assessments for learners of English for Specific Purposes (ESP) entails distinctive methodological and pedagogical considerations that extend beyond the framework of general English testing. Unlike general proficiency examinations, ESP reading assessments must integrate domain-specific discourse, specialized vocabulary, and authentic task design. This article examines the peculiarities inherent in the development of ESP reading assessments, drawing upon illustrative examples from the fields of law, business, international communication, the study identifies challenges related to authenticity, lexical density, and task design, while also providing evidence-based recommendations for test developers.

Introduction. English for Specific Purposes (ESP) represents a specialized branch of language learning that seeks to address the communicative and linguistic requirements of learners operating within defined academic, professional, or vocational contexts. ESP, as defined by Hutchinson and Waters (1987), prioritizes learners' professional or academic needs, situating language learning within specific disciplinary or vocational contexts. In this regard, reading occupies a central position, as it constitutes a primary means of accessing professional knowledge, engaging with academic research, and performing workplace tasks. A central component of ESP instruction is the development of learners' ability to comprehend, interpret, and critically engage with texts specific to their disciplinary or occupational domain.

Consequently, reading assessments within ESP curricula assume a critical role in evaluating learners' progress, preparedness, and capacity to function effectively in target domains.

The design of ESP reading tests is inherently more complex than that of general English assessments. General proficiency tests typically emphasize topics of universal familiarity, whereas ESP assessments require alignment with specialized discourse and professional communicative practices. For instance, a business English learner may be required to interpret a company's annual financial report, while a law student must extract relevant provisions from a contractual clause. In the diplomatic sphere, a learner may be tasked with interpreting a passage from a United Nations resolution or a policy communiqué, while in international communication, the focus may lie on analyzing press releases or media coverage of intercultural negotiations. As Bachman and Palmer (1996) argue, test usefulness depends on validity, reliability, authenticity, interactiveness, and practicality. In the ESP context, authenticity and domain-specificity assume particular importance, yet they must be balanced with fairness and accessibility for learners whose professional expertise may vary. Weir (2005) further emphasizes the importance of cognitive validity, highlighting the need to align test tasks with real-life reading processes in professional settings.

The specificities of ESP test development thus derive from the necessity of combining linguistic competence with domain knowledge, without allowing the latter to overshadow the primary goal of language assessment.

Methods. The methodological approach for this paper draws upon a qualitative analysis of established testing frameworks and applied practices in ESP assessment. Key theoretical perspectives were drawn from Bachman (1990), Bachman and Palmer (1996), Alderson (2000), and Weir (2005), whose works provide foundational guidance on language testing principles. Complementary insights were obtained from ESP literature, including Dudley-Evans and St John (1998), Douglas (2000), and Hutchinson and Waters (1987). To contextualize these principles, authentic examples from multiple professional domains were analyzed and the following stages were implemented:

The process of ESP reading test development begins with a comprehensive needs analysis, which involves identifying target learner groups such as law students, medical students, business professionals, engineers, diplomats, and communication specialists. For each of these groups, typical reading practices in authentic professional contexts must be carefully delineated in order to ensure that assessment tasks accurately reflect real-world communicative demands.

Following this stage, test specifications are formulated to provide a structured blueprint for assessment design. These specifications define the relevant reading sub-skills to be evaluated—including skimming, scanning, inferencing, and critical reading—while also determining suitable item formats such as multiple-choice, matching, short-answer, true/false, and open-ended questions.

The next stage involves task design, which requires the careful selection and compilation of authentic reading passages from domain-relevant sources. Illustrative examples include contractual provisions in the legal field, financial disclosures in business, diplomatic notes and multilateral agreements, as well as institutional press releases in international communication. These materials are adapted as necessary to ensure accessibility while preserving the authenticity of professional discourse.

Pilot testing plays a crucial role in refining assessment instruments. Prototype tests are administered to representative groups of ESP learners in order to evaluate task clarity, cognitive load, difficulty level, and contextual relevance. Feedback obtained during this stage provides essential insights for subsequent revisions.

Finally, the analysis of reliability and validity ensures that the assessment instruments are both consistent and accurate in measuring the intended construct. Reliability is defined as the stability of results across different administrations, while validity pertains to the extent to which the test measures the specific reading competencies it purports to assess. In addition, practical considerations—including time constraints, resource availability, and scoring feasibility—are evaluated to guarantee that the assessment is not only theoretically sound but also implementable in real-world educational contexts. The methodological design integrates both conceptual exploration and practical illustration, aiming to bridge theoretical frameworks with authentic professional applications. The process of ESP reading test development was organized into several interrelated stages, ensuring that the assessment instruments not only align with language testing principles but also meet the communicative needs of specific learner groups.

ESP reading test development differs from general English testing primarily in its situated specificity and context-dependence. Unlike general proficiency tests, ESP assessments must mirror the textual genres, discourse conventions, and cognitive operations inherent to each professional field. Consequently, the test designer must navigate the balance between authenticity and accessibility, ensuring that domain-specific texts remain pedagogically manageable while retaining professional realism. This dynamic process underscores the interdisciplinary nature of ESP test development, where linguistic competence is assessed not

in isolation but as an instrument of professional communication. The methodological approach outlined in this study demonstrates how theoretical principles and practical constraints converge to create robust, valid, and contextually meaningful reading assessments for ESP learners.

Results. The application of the aforementioned methodology produced several salient findings regarding the peculiarities of ESP reading test development.

A central finding of the study concerned the importance of content authenticity, which emerged as a key determinant of test validity. Authenticity has long been emphasized as a cornerstone of language assessment (Bachman & Palmer, 1996; Weir, 2005), yet texts drawn directly from professional contexts often exceeded the linguistic and conceptual proficiency of learners. To address this challenge, careful selection and adaptation of materials proved essential. For example, in the legal domain, an excerpt from a confidentiality clause was used to assess comprehension through a task such as: “According to the text, under what conditions may confidential information be disclosed?” In diplomacy, a paragraph from a United Nations Security Council resolution served as stimulus material, accompanied by the item: “What specific action does the resolution request of member states?” Both examples reflected the interpretive practices expected of novice professionals in their respective fields while remaining accessible to language learners.

Another salient issue involved the lexical density of ESP texts, which are often characterized by specialized terminology. As Alderson (2000) notes, vocabulary load is a critical factor in determining text difficulty and, consequently, test fairness. Medical case descriptions, for instance, frequently contain terms such as idiopathic, etiology, or pathogenesis, while diplomatic texts regularly employ institutionalized expressions such as confidence-building measures, territorial integrity, or memorandum of understanding. Although such terms must be preserved for authenticity, test items should not require extensive prior subject-matter knowledge. In medicine, for example, learners were presented with a short clinical note and asked to match specialized terms with plain-language equivalents. In the field of international communication, a press release containing institutional terminology required learners to answer a task such as: “Which of the following phrases refers to a non-binding agreement?” These strategies align with Weir’s (2005) notion of linguistic accessibility, ensuring that comprehension—not prior expertise—is being assessed.

The authenticity of tasks themselves was identified as a crucial factor in effective test design. Task authenticity, understood as the degree to which an assessment mirrors real-world language use (Douglas, 2000), proved essential in fostering meaningful engagement. Business

students, for instance, were presented with a balance sheet and asked: “Which figure represents the company’s liabilities?” Engineering students encountered excerpts from technical user manuals, with tasks such as: “What is the maximum recommended temperature to prevent overheating?” Similarly, diplomatic learners were tested with communiqués from peace negotiations, where the task required them to identify principles of international law emphasized in key phrases such as territorial integrity. In international communication, learners analyzed journalistic texts covering diplomatic crises and were asked: “Which evaluative stance does the journalist adopt toward the minister’s statement: critical, neutral, or supportive?” These examples highlight the importance of designing tasks that reflect professional literacy practices while remaining pedagogically appropriate.

In addition, the integration of skills was observed to be a defining characteristic of ESP reading assessments. Learners were frequently required not only to comprehend textual information but also to apply it to contextually relevant scenarios, echoing the construct-focused principles of validity highlighted by Bachman (1990) and Alderson (2000). Law students, for example, read case judgments and responded to prompts such as: “Which contractual clause was violated, and how did the court interpret it?” Similarly, diplomatic learners studied excerpts from the Vienna Convention and answered interpretive questions such as: “What obligations does a diplomatic mission owe to the host state, as stipulated in the text?” These integrated tasks extended beyond surface-level comprehension, requiring learners to combine reading, interpretation, and application skills.

Finally, issues of reliability and practicality played a significant role in shaping test development. Pilot testing, as recommended in test validation frameworks (Alderson, Clapham, & Wall, 1995; Weir, 2005), revealed that assessments overly dependent on domain-specific knowledge risked disadvantaging learners with limited subject expertise. Engineering students unfamiliar with advanced thermodynamic principles or diplomacy students unacquainted with specific United Nations protocols often struggled, not because of linguistic shortcomings, but because of gaps in their disciplinary background. To mitigate this issue, materials were carefully adapted to ensure accessibility while maintaining a high degree of authenticity. This balance between professional relevance and linguistic fairness proved indispensable for creating valid and reliable ESP reading assessments.

Discussion. The findings of the present analysis underscore the inherent complexity of developing ESP reading assessments, particularly due to the necessity of ensuring authenticity, balance, and fairness. One of the central challenges involves establishing an equilibrium between linguistic competence and domain-specific expertise. ESP reading assessments must

prioritize the evaluation of comprehension skills rather than the demonstration of professional knowledge. For instance, medical students should be assessed on their ability to interpret case reports in English, rather than on their capacity to provide accurate diagnoses. Similarly, learners specializing in diplomacy should be required to demonstrate comprehension of communiqués or multilateral agreements, rather than advanced mastery of international legal frameworks.

Authenticity constitutes another cornerstone of validity in ESP assessment (Bachman & Palmer, 1996). However, unmodified professional texts frequently impose excessive cognitive and lexical demands on learners. Adaptation strategies, such as selective excerpting, paraphrasing, and the reduction of jargon density, enable test developers to preserve the realism of professional discourse while maintaining accessibility. For example, a United Nations General Assembly address may be reduced to essential paragraphs without compromising its rhetorical and thematic integrity. Such modifications allow learners to engage with authentic genres without being overwhelmed by their linguistic or conceptual complexity.

A further consideration is the use of varied task formats. The inclusion of multiple item types not only enhances construct validity but also mitigates test fatigue and sustains learner engagement. Formats such as multiple-choice questions, gap-filling, matching, and short-answer tasks allow for the assessment of different reading sub-skills, including skimming, scanning, inferencing, and critical reading. Importantly, these task types must reflect authentic professional practices. For instance, tasks may require the extraction of key arguments, the comparison of data sets, the interpretation of procedural instructions, or the evaluation of authorial stance and rhetorical tone. By aligning task design with actual communicative practices, test developers ensure that assessments remain both pedagogically meaningful and professionally relevant.

Equally important is the issue of domain-specific relevance. Effective ESP reading assessment necessitates the use of materials that closely correspond to the communicative demands of specific fields. In legal contexts, these may include statutes, contractual clauses, or judicial opinions; in business, annual financial reports or marketing case studies; in medicine, clinical case histories and research abstracts; in engineering, technical manuals and design specifications; and in diplomacy, treaties, communiqués, and official press releases. The careful alignment of test content with professional genres ensures ecological validity and better prepares learners for the communicative realities of their respective domains.

The principle of fairness is also paramount. As Alderson (2000) emphasizes, fairness is an indispensable criterion in assessment. ESP reading tests must be designed in such a way that

they do not confer undue advantage upon candidates with advanced subject-matter expertise. The focus must remain on the evaluation of linguistic and interpretive skills rather than professional knowledge. Ensuring this balance is particularly important in disciplines where technical or institutionalized language may otherwise obscure the linguistic construct being measured.

Thus, ESP test development must be understood as an iterative process, requiring continuous revision and refinement. Pilot testing, expert review, and learner feedback are essential for evaluating clarity, difficulty level, and contextual relevance. Weir's (2005) model of test validation further underscores the necessity of ongoing revision to maintain validity and reliability. Moreover, the dynamic nature of professional discourse—particularly in diplomacy and international communication, where geopolitical developments evolve rapidly—necessitates the regular updating of test materials to preserve authenticity and contextual appropriateness. Through this process of continuous adaptation, ESP reading assessments can remain both rigorous and responsive to the evolving communicative needs of learners.

Test specification for ESP reading assessments

Domain	Target texts	Reading skills assessed	Task types	Examples
Law	Contracts, statutes, case summaries	Skimming for legal provisions, inferring implications	Multiple-choice, matching, short-answer	Match contractual clauses with their legal effects.
Business	Annual reports, market analyses	Identifying key data, evaluating arguments	Chart interpretation, gap-fill, summary	Extract financial trends from a company report.
Diplomacy	Treaties, communiqués, press releases	Identifying stance, tone, and key provisions	Multiple-choice, summary completion, MCQs	Identify the central position expressed in a diplomatic note.

Conclusion. The present study has demonstrated that the development of reading assessments in the field of ESP constitutes a qualitatively distinct process from the design of general English proficiency tests. As emphasized by Bachman (1990) and Bachman and Palmer

(1996), effective assessment design must align with both the communicative purpose and the contextual features of language use. In the ESP context, this alignment necessitates the incorporation of authentic disciplinary discourse, specialized lexical and syntactic patterns, and professionally relevant task types, while preserving linguistic accessibility and assessment fairness. The findings of this study suggest that the peculiarities of ESP reading assessment lie in balancing authenticity with comprehensibility, integrating multiple reading strategies, and ensuring that test performance reflects language ability rather than domain-specific expertise—a distinction underscored by Douglas (2000) in his discussion of specific purpose language testing. Consequently, effective ESP reading assessments must fulfill the essential criteria of validity, reliability, and practicality as articulated in Weir's (2005) socio-cognitive framework, while remaining closely aligned with learners' professional communicative requirements. The methodological processes of needs analysis, careful task design, pilot testing, and iterative refinement - recommended by Alderson (2000) and Dudley-Evans and St John (1998) - are therefore indispensable for ensuring both theoretical coherence and contextual appropriateness.

The significance of such assessments is particularly evident in high-stakes professional contexts, such as international communication and diplomacy, where textual misinterpretation may entail serious professional, institutional, or even political consequences. Within these domains, ESP reading assessments contribute not only to the linguistic development of learners but also to the broader objective of promoting accurate, contextually sensitive, and ethically responsible communication across linguistic and cultural boundaries (Hutchinson & Waters, 1987; Douglas, 2000). Thus, in line with current trends toward evidence-based assessment practices and ongoing validation research, it is essential that test developers and educators adopt a reflective, adaptive approach to ESP test design. By maintaining a commitment to continuous revision and empirical validation, ESP reading assessments can evolve alongside the dynamic communicative demands of specialized professional environments. In doing so, they contribute to the preparation of professionals who are not only linguistically proficient but also strategically competent and interculturally aware, capable of engaging effectively with the complex textual and pragmatic practices of their disciplines.

References:

1. Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
2. Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.

3. Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford University Press.
4. Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.
5. Douglas, D. (2000). Assessing languages for specific purposes. Cambridge University Press.
6. Umida A. Stylistic Peculiarities of Herbert Bates Stories //International journal of inclusive and sustainable education. – 2023. – T. 2. – №. 2. – C. 122-125.
7. Yusupovna A. U., Shavkatovna I. M. Development of Reading Skills on the Basis of Synthetic Method in a Non-Linguistic Universities //American Journal of Language, Literacy and Learning in STEM Education (2993-2769). – 2023. – T. 1. – №. 9. – C. 241-244.