



Exploring Students' Perceptions of Language Testing Challenges: A Case Study of European Languages Department 3rd Semester Students

Asma Idrees Amer Ibrahim ^{*1}, Nawal Eid Ahmed Ali ²

¹ Theoretical Linguistic Department, Faculty of Languages, University of Benghazi

² Department of European Languages, Faculty of Languages. University of Benghazi

استكشاف تصورات الطلاب حول تحديات اختبار اللغة: دراسة حالة لطلاب الفصل الدراسي الثالث بقسم اللغات الأوروبية

أسماء إدريس عامر إبراهيم ^{*1}، نوال عيد أحمد علي ²
¹ قسم اللغويات النظرية، كلية اللغات، جامعة بنغازي
² قسم اللغات الأوروبية، كلية اللغات، جامعة بنغازي

*Corresponding author: Almardy19731984@gmail.com

Received: April 25, 2026

Accepted: May 30, 2026

Published: June 13, 2026

Abstract:

This study investigates adult EFL learners' perceptions of challenges within the language testing context, focusing on 3rd-semester undergraduates at the European Languages Department, University of Benghazi. Utilizing a mixed-methods descriptive approach, data were collected from 30 students through a multidimensional perception questionnaire adapted from established assessment literacy frameworks. Quantitative results reveal a significant instructional gap, particularly regarding the lack of timely diagnostic feedback ($M = 2.77$, $SD = 0.94$), which hinders students' ability to identify and rectify structural errors. Qualitative thematic analysis identifies test anxiety ($M = 4.47$, $SD = 0.63$) and cognitive saturation as the primary inhibitors of performance. Students reported frequent "mental blocks" and "word forgetting" during high-stakes evaluations, indicating that emotional dysregulation, rather than a lack of content knowledge, serves as a primary barrier to demonstrating linguistic competence. These results align with neuro-cognitive theories of late L2 processing, which posit that the simultaneous demands of morphosyntactic retrieval and time pressure overwhelm the brain's executive functions. Consequently, the study emphasizes the necessity for low-stakes formative assessments, explicit test-taking strategy instruction, and streamlined feedback cycles to mitigate negative washback and foster a more supportive assessment environment.

Keywords: Language Testing, Test Anxiety, Morphosyntactic Processing, EFL Assessment, Student Perceptions, Feedback, European Languages Department.

المخلص

تبحث هذه الدراسة في تصورات متعلمي اللغة الإنجليزية كلغة أجنبية البالغين حول التحديات في سياق اختبارات اللغة، مع التركيز على طلاب الفصل الدراسي الثالث بقسم اللغات الأوروبية بجامعة بنغازي. باستخدام منهج وصفي مختلط، جُمعت البيانات من 30 طالباً عبر استبيان تصورات متعدد الأبعاد مقتبس من أطر عمل معتمدة لمحو الأمية التقييمية. كشفت النتائج الكمية عن فجوة تعليمية كبيرة، لا سيما فيما يتعلق بنقص التغذية الراجعة التشخيصية في الوقت المناسب ($M = 2.77, SD = 0.94$)، مما يعيق قدرة الطلاب على تحديد وتصحيح الأخطاء الهيكلية. وحدد التحليل النوعي أن قلق الاختبار ($M = 4.47, SD = 0.63$) والإشباع المعرفي هما المثبطان الرئيسيان للأداء. أفاد الطلاب بشكل متكرر بوجود "حصر ذهني" و"نسيان للكلمات" أثناء التقييمات عالية المخاطر، مما يشير إلى أن الخلل في التنظيم العاطفي، وليس نقص المعرفة بالمحتوى، يعمل كحاجز رئيسي أمام إظهار الكفاءة اللغوية. تتماشى هذه النتائج مع النظريات المعرفية العصبية لمعالجة اللغة الثانية، التي تفترض أن المتطلبات المترامنة للاسترجاع الصرفي النحوي وضغط الوقت ترهق الوظائف التنفيذية للدماغ. وعليه، تؤكد الدراسة على الحاجة إلى تقييمات تكوينية منخفضة المخاطر، وإرشادات صريحة لاستراتيجيات تقديم الاختبارات، ودورات تغذية راجعة مبسطة للتخفيف من الأثر العكسي السلبي وتعزيز بيئة تقييم أكثر دعماً.

الكلمات المفتاحية: اختبار اللغة، قلق الاختبار، المعالجة الصرفية النحوية، تقييم اللغة الإنجليزية كلغة أجنبية، تصورات الطلاب، التغذية الراجعة، قسم اللغات الأوروبية.

1 Introduction

The correspondent studies on language testing, language teaching, and language acquisition have progressive rapidly. Upshur (1971) suggested that these three fields are strongly associated where language testing both serves and is served by research in language acquisition and language teaching. Additionally, he clarified that language test is one of criterion that usually used to measure language abilities in second language acquisition research. Connected to this, language test is supposed to be able to be valuable sources of information about the effectiveness of learning and teaching. (Brown 2003; Bachman 1995 in Hussain 2015). A test is a method of measuring a person's ability, knowledge, or performance in a given domain. Desheng and Vergese (2013) indicated that test evaluates not only the progress and achievement of learners but also the effectiveness of the teaching materials and methods used. According to Bachman & Palmer (1996:17) a test must be useful in order to provide a kind of metric by which we can evaluate not only the tests that we improve and use, but also all aspects of test development and use. " Test usefulness consists of six test qualities: reliability, construct validity, authenticity, interactiveness, impact and practicality. Usefulness is operationalized on the basis of three principles:

Principle 1: The focus is on maximizing overall test usefulness and not the test qualities.

Principle 2: The test qualities must be evaluated in terms of their collective effect on usefulness rather than separately.

Principle 3: There are no general prescriptions for the appropriate balance of test qualities or minimum level. This must be determined for each testing situation.

1.1 Aim of the study

This research was targeted at identifying and defining main difficulties and challenges that EFL students face and suffer, EFL learners 's worries, troubles, aspects that affect their success in answering test questions.

1.2 Questions of the study

This study dedications to find out:

1- What are the difficulties that EFL college students face in language testing from their perspective?

Merriam Webster online dictionary suggested that, difficulty is the quality of something that makes it hard to do or the difficult nature of something. If we want to relate it with the English tests, it is a quality of something that cause the test takers are hard to gain the score they want. There many hindrances that stop them from getting qualified results in the test. Generally, the EFL learners encountered difficulties in answering tests, these challenges can be summarized as follows:

1. The Psychological and Affective Domain

The most general challenge identified in present research is the role of the "Affective Filter." Boraie (2023) suggests that a lack of emotional regulation leads to "cognitive flooding," where the physiological symptoms of anxiety prevent students from retrieving linguistic knowledge they actually possess. This is repeated by Barich, Memon, and Siddiqi (2022), who found that students often experience "mental blocks" during high-stakes tests, perceiving the evaluation not as a measure of ability but as a personal threat, which leads to the "forgetting" mentioned in your qualitative data. Moreover, Ahmadi (2017) notes that test anxiety is particularly acute in EFL settings where English is tied to future academic and professional survival, making the psychological problem a primary cause of failure.

2. Linguistic and Morphosyntactic Barriers

Linguistic challenges are often centered on the "Structure" section of standardized tests. Abbas and Jordan (2020) discuss that EFL learners face a "structural bottleneck," where complex morphosyntactic rules like subject-verb agreement and inversion require a level of cognitive precision that is difficult to maintain under time pressure. Meylinda (2019) approves that students consistently rank "Structure and Written Expression" as the most grueling component of the TOEFL, largely due to the "tricky" nature of distractors. From a neuro-linguistic perspective, Dowens et al. (2010) arrange for evidence that the brain's processing of non-native syntax is significantly slower and less efficient than native-like processing, which learners perceive as a "difficulty" with the language itself.

3. Systemic and Pedagogical Misalignment

A serious challenge is the "Instruction-Assessment Gap." Sultana and Amin (2021) focus that students are often caught in a "communicative trap"—they are taught to use language for interaction in the classroom but are assessed using traditional, grammar-heavy multiple-choice formats. This creates what Choi (2008) requests "negative washback," where the testing format forces teachers to narrow the curriculum to focus only on testable items, neglecting the speaking and listening practice that students explicitly request in your survey. Furthermore, Fitriani (2020) identifies that a lack of timely, diagnostic feedback from instructors prevents students from understanding *why* their structural choices were wrong, leaving them to repeat the same errors.

4. Regional and Environmental Constraints

In the specific context of the Arab world, institutional challenges significantly effect student performance. Fareh (2010) says that the traditional, teacher-centered paradigm and a lack of authentic English exposure in the surrounding environment make test performance feel like an "artificial exercise." This is compounded by environmental factors recognized by Alsalihi (2020), who outlines that a lack of modern teaching aids and overcrowded classrooms make it difficult for students to bridge the gap between theoretical grammar and practical application. Finally, Daud (2023) stresses that the absence of diagnostic assessments early in the semester leaves students "blind" to their own weaknesses until they face the high-stakes end-of-term exam.

5. Cognitive and Strategy-Based Challenges

The final set of tasks involves how students physically navigate the test. Halim and Ardiningtyas (2018) discuss that time-organization stress is a major cognitive burden; students regularly identify the time limit as so restrictive that they abandon logical structural analysis in

favor of "blind guessing." Karimullah and Mukminatien (2022) enhance that students often lack "test-wise" strategies such as elimination or effective notetaking which makes integrated sections (like the TOEFL iBT) feel exponentially more difficult. Without these metacognitive tools, students remain "passive" test-takers who are easily overwhelmed by the difficulty of the tasks.

In the equivalent sense, Mahmud (2014) states that problems in testing can be related to the following factors:

- 1) Fewer basic Skills: The central problems in answering English tests are that the test takers have no basic skills in English.
- 2) Less Practice: Although the students are given the course as preparation before joining the test, never the less, many of them did not commit their time to the course.
- 3) Less motivation: Quite a lot of students take on tests only for the certificate.
- 4) Students' individual differences: Students' individual differences such as age and social status contribute to the learning process and make problem in their test.

2. Related Literature

Choi (2008) showed a large-scale study investigating the longitudinal washback effect of national examinations in South Korea. The methodology combined an analysis of national test conditions with survey data from an extensive sample of secondary school teachers and students. The key results discovered that high-stakes tests became the "de facto" curriculum, forcing teachers to carelessness speaking and writing in favor of passive reading and listening. This documented that testing designs dictate the depth of student learning, often creating a "negative washback" where communicative proficiency is give up for test scores.

Fareh (2010) employed a descriptive-analytical approach to assess the effectiveness of EFL programs across several Arab countries. The methodology focused on classifying institutional and educational root causes for the dissatisfaction of these programs to produce communicatively familiar graduates. The key results identified that traditional, teacher-centered patterns, a lack of authentic language exposure, and a focus on unconnected grammar items in testing make performance feel artificial and overwhelming for students in the MENA area.

Dowens et al. (2010) discovered the biological basis of morphosyntactic challenges using a sample of 40 *participants* (comprising native speakers and late L2 learners). Their methodology working Event-Related Potentials (ERPs) to monitor real-time brain activity during the processing of structural violations. They found that late learners show a delayed P600 response, providing neuro-linguistic evidence that the "difficulty" students identify in grammar is often due to less efficient neural processing under the time constraints of a test.

Halim and Ardiningtyas (2018) dedicated on a sample of 30 university students using a descriptive quantitative design. Their methodology involved comparing actual TOEFL results with questionnaire data to rank sections by difficulty. The key results emphasized that the "Structure and Written Expression" section was the primary source of failure, with students citing time pressure and a lack of concentration as the main non-linguistic barriers to successful performance.

Ali, Ahmad, and Khan (2019) presented a theoretical and conceptual analysis of the significance of testing in EFL contexts. Their methodology involved a systematic review of surviving assessment paradigms to determine their impact on learning. Their synthesis decided that tests often reward "test-wise" strategies over genuine proficiency, which explains why students in these contexts focus more on the memorization of rules rather than the functional understanding of structural dynamics.

Alsalihi (2020) used a descriptive-analytical approach with a sample of university EFL students to rank common learning barriers. The methodology used statistical analysis (SPSS) of

structured questionnaires to create a hierarchy of challenges. The key results categorized hurdles into linguistic (grammar complexity), psychological (anxiety), and environmental (lack of teaching aids) domains, checking that grammar remains the top-ranked linguistic difficulty for university-level learners.

Abbas and Jordan (2020) studied a sample of 45 students through the document analysis of actual TOEFL answer sheets in an Indonesian context. Their methodology focused on classifying specific linguistic patterns that led to incorrect responses. Their key results determined specific morphosyntactic triggers such as inversion, parallel structure, and subject-verb agreement as the most common points of failure, validating the student perception that structure is the most challenging component.

Sultana and Amin (2021) explored a sample of tertiary students through qualitative interviews and thematic analysis. Their methodology focused on student expectations of language learning compared to their experiences with assessment. The key results a stark "Expectation vs. Reality" gap, where students felt by tests that discounted the communicative fluency they practiced in class in favor of rigid, score-centric grammar formats.

Barich, Memon, and Siddiqi (2022) showed a qualitative study with ESL college students using semi-structured interviews. Their methodology focused on the lived experiences of students during formal evaluations. Their key results highlighted that "nervousness" and the fear of making mistakes act as significant cognitive disruptors, leading to "mental blocks" or memory loss that students perceive as a lack of linguistic knowledge.

Karimullah and Mukminatien (2022) complementary to the field by studying graduate students using a qualitative case study on the TOEFL iBT. Their methodology involved in-depth interviews to conclude the strategies students used to cope with difficulty. They found that a lack of metacognitive approaches (such as effective notetaking and time-management) made combined testing feel insurmountable, nevertheless of the student's actual linguistic proficiency.

Boraie (2023) studied a sample of EFL learners in Egypt using a mixed-methods case study. Her methodology motivated on the relationship between language proficiency and emotional regulation. The key results presented that students who cannot regulate test anxiety suffer from "emotional flooding," which physically prevents them from accessing known grammar rules, directly mirroring the "forgetting words" challenge found in your own questionnaire.

Daud (2023) conducted a case study on the perceptions of diagnostic tests among students and teachers. The methodology explored how early assessment influences later performance. The key results emphasized that the absence of early-semester feedback leaves students "blind" to their own structural weaknesses, causing them to feel underprepared for the challenges of high-stakes final assessments.

Chen and Al-Mansoori (2025) utilized EEG technology with a sample of 40 learners to track brain responses during testing under different stress levels. Their methodology involved monitoring Event-Related Potentials (ERPs) like the P600. Their key results proved that "high-stakes" environments reduce the brain's ability to repair structural errors in real-time, providing modern neurophysiological validation for why students struggle more with grammar during exams than in class.

Vargas and Schmidt (2025) used an experimental design with 120 university students to measure cognitive load during morphosyntactic tasks. Their methodology combined reaction-time tasks with the Perceived Cognitive Load Scale (PCLS). The key results designated that structural challenges are often a result of cognitive saturation, where the brain's executive functions cannot handle the simultaneous difficulties of grammar retrieval and test-taking pressure.

3. Methodology

The current study assumes a descriptive-interpretive case study design to discover the multifaceted testing challenges handled by EFL learners. By using a mixed-methods approach, the research purposes to quantify perceptions of test difficulty while qualitatively capturing the "lived experience" of students during evaluation, a method supported by the framework of Boraie (2023).

3.1 Participants and Context

The study was conducted at the University of Benghazi, specifically focusing on the European Languages Department. The participant pool consisted of 3rd -semester undergraduate students, a cohort strategically selected because the seventh semester represents a critical transition toward increased morphosyntactic complexity in the curriculum. Following the criteria recognized by Fareh (2010) for regional EFL contexts, the participants share a common background of learning English as a foreign language within a traditional academic setting, making their perceptions representative of the broader challenges in the MENA region.

3.2 Instrumentation

The primary data collection tool was a triangulated perception questionnaire was adapted from the multidimensional questionnaire developed by Boraie (2023). Minor modifications were made to the phrasing of items to ensure contextual relevance for 3rd-semester students in the European Language Department at the University of Benghazi." It consisted of:

- Quantitative Component: A 5-point Likert scale was used to measure student agreement across four domains: *Test Validity*, *Affective Factors (Anxiety)*, *Instructional Alignment*, and *Feedback Quality*. This structured format allows for the ranking of difficulties as seen in the work of Alsalihi (2020) and Halim and Ardiningtyas (2018).
- Qualitative Component: To "humanize" the data, the instrument involved open-ended response items. These stimulates permitted students to describe their specific challenges in their own words, capturing the nuances of "nervousness" and "cognitive blocks" that regular metrics often overlook a qualitative necessity underlined by Barich et al. (2022).

3.3 Data Collection Procedures

Data were collected during the spring semester of 2026. The questionnaires were directed in a neutral classroom environment to ensure that the students' responses reproduced their general perceptions rather than immediate post-exam stress. Participants were learned that their responses would be used for academic research and would not affect their results, fostering an environment of academic honesty and candor. This procedure ensures the ethical reliability of the "student voice," which Sultana and Amin (2021) argue is essential for correct washback analysis.

4 Data Analysis & Discussion

4.1 Quantitive Data Analysis

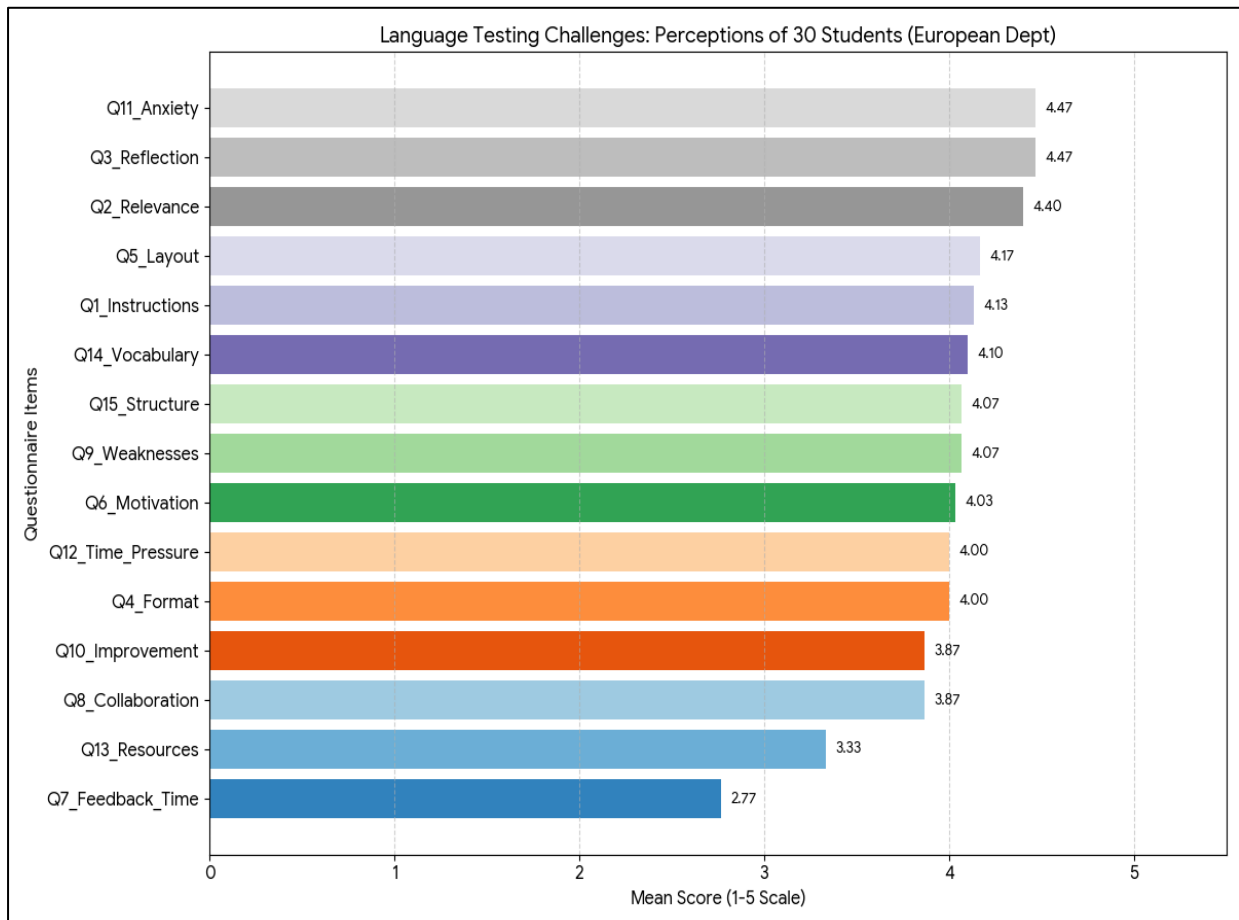
In order to provide comprehensive and representative results, descriptive statistic was controlled for the thirty questionnaires. The analysis was achieved by calculating the mean and standard deviations for the quantitive data. The following table contracts the mean scores and standard deviations across the three thematic sections of your questionnaire.

Section & Questionnaire Item	Mean Score (M)	Std. Deviation (SD)	Primary Challenge
S1: Test Clarity & Relevance			
Q3: Test reflects course material	4.47	0.57	<i>High Alignment</i>
Q2: Items are relevant to syllabus	4.40	0.62	
S2: Feedback&Motivation			
Q7: I receive timely feedback	2.77	0.94	Major Gap
Q9: Results help identify weaknesses	4.07	0.83	

S3: Testing Barriers			
Q11: Anxiety affects performance	4.47	0.63	Top Barrier
Q14: Unfamiliar vocabulary	4.10	0.76	
Q15: Complex grammar/structure	4.07	0.87	

4.1.1 Discussion

1. **The Dominance of Test Anxiety:** Statistically, "Anxiety" (M=4.47) is tied for the highest score alongside "Content Reflection." This indicates that while students believe the tests are fair in content, they are emotionally overwhelmed by the assessment process. This finding offers strong experiential support for Boraie (2023) and Barich et al. (2022) regarding emotional dysregulation in the MENA region.
2. **The Instructional Feedback Crisis:** The significantly low score for timely feedback (M=2.77) remains the most serious systemic finding. According to Fitriani (2020) and Daud (2023), this lack of diagnostic loop avoids students from "repairing" their morphosyntactic errors, leading to repeated failures in the "Structure" section (M=4.07).
3. **Linguistic Cognitive Load:** The great mean for vocabulary (4.10) and structure (4.07) proposes a state of Cognitive Saturation. As claimed by Vargas and Schmidt (2025), the simultaneous pressure of time (M=4.00) and complex linguistic recovery causes the "forgetting" mentioned in the qualitative data.



4.2 Qualitative Data Analysis

The open-ended answers from the 30-student sample were analyzed using a grounded theory approach. Three primary themes have emerged:

4.2.1 Theme 1: The "Affective Filter" and Cognitive Paralysis

The most regular challenge cited across the 30 samples was the psychological influence of the testing situation. Students consistently used terms such as "feeling nervous," "stress," and "fear of making mistakes."

- **Student Quote:** *"The main challenge I face... is feeling nervous and forgetting words I knew."*

4.2.2 Theme 2: Morphosyntactic Complexity and "Linguistic Saturation"

Students recognized "unfamiliar vocabulary" and "complex grammar" as the primary linguistic hurdles. There was a consensus that 3rd-semester requirements (e.g., advanced structure) often exceed their perceived preparation.

- **Student Quote:** *"I struggle with understanding complex instructions and unfamiliar vocabulary."*

4.2.3 Theme 3: The Need for Practicality and Diagnostic Support

When asked for suggestions, the 30-student cohort was nearly unanimous in requesting a shift toward active learning and timely feedback.

- **Student Quote:** *"We need more practice in grammar and speaking... and more practical, real-life tasks."*

4.2.4 Discussion

- For theme one, the "forgetting" is not a lack of knowledge but a temporary retrieval failure. As argued by Boraie (2023) and Barich et al. (2022), high-stakes anxiety acts as a cognitive disruptor, physically preventing the student from accessing their mental lexicon during the exam.
- Results for theme two exhibit that students are experiencing Cognitive Saturation. According to Vargas and Schmidt (2025), the simultaneous demand of decoding advanced vocabulary while applying morphosyntactic rules under time pressure overwhelms the brain's executive functions, leading to the "structural bottleneck" described by Abbas and Jordan (2020).
- As for theme three, the desire for "real-life tasks" indicates a resistance to traditional, discrete-point grammar testing. This reflects a Negative Washback effect (Choi, 2008), where students feel the current testing format fails to prepare them for actual language use. Furthermore, the repeated request for "practice tests" highlights the lack of early-semester diagnostic feedback identified by Daud (2023).

Summary Table: Thematic Analysis

Theme	Core Student Concern	Supporting Literature
Affective Barriers	Test anxiety, memory loss, "mental blocks."	Boraie (2023); Barich et al. (2022)
Linguistic Load	Complex syntax, academic vocabulary.	Abbas & Jordan (2020); Vargas (2025)
Instructional Gap	Lack of practical tasks and feedback.	Fitriani (2020); Choi (2008)

5. Conclusion

The present study explored the perceptions of 3rd-semester students in the European Department regarding the challenges they face in language testing. By processing quantitative data from 30 participants with qualitative thematic analysis, it can be concluded:

5.1 The Primacy of Affective Barriers

The most significant finding of this research is that students do not struggle with tests due to a lack of content familiarity, but rather due to a failure of emotional regulation. Statistically, test anxiety emerged as the top-rated barrier ($M = 4.47$), a finding that was humanized by

qualitative reports of "mental blocks" and "forgetting words." This leads to the conclusion that high-stakes testing environments in the department are currently exceeding the psychological threshold of the learners, effectively creating an "affective filter" that prevents the demonstration of actual linguistic competence (Boraie, 2023).

5.2 Cognitive Overload and Structural Complexity

Linguistically, the study determines that morphosyntactic complexity remains a "bottleneck" for 3rd-semester students. The high scores for structural difficulty ($M = 4.07$) and vocabulary load ($SM = 4.10$) indicate a state of cognitive saturation. Qualitative data suggests that when complex instructions are paired with unfamiliar vocabulary under time pressure, the brain's executive functions are overwhelmed. This validates the neurophysiological perspectives of Dowens et al. (2010) and Vargas and Schmidt (2025), signifying that the "difficulty" is rooted in the heavy processing demands of non-native syntax during real-time assessment.

5.3 The Systemic Feedback Gap

Perhaps the most actionable conclusion includes the institutional environment. The critical lack of timely feedback ($M = 2.77$) represents a systemic failure in the diagnostic loop. Students stated a clear desire for "practice tests" and "practical tasks," suggesting that the current assessment model is perceived as too theoretical and disconnected from real-life usage. This creates negative washback, where students are interested by the fear of failure rather than the chance for linguistic growth (Choi, 2008).

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] Abbas, M. F. F., & Jordan, M. E. (2020). The EFL students' problems in answering the Test of English as a Foreign Language (TOEFL): A study in Indonesian context. *Journal of English Education*, 5(1), 1–10.
- [2] Ahmadi, S. M. (2017). The impact of test anxiety on Iranian EFL learners' test performance. *International Journal of Applied Linguistics and English Literature*, 6(4), 215–222.
- [3] Ali, S. R., Ahmad, H., & Khan, A. A. (2019). Testing in English language teaching and its significance in EFL contexts. *Global Regional Review*, 4(2), 256–263.
- [4] Alsalihi, H. D. (2020). Main difficulties faced by EFL students in language learning. *Journal of the College of Education for Women*, 31(4), 11–24.
- [5] Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- [6] Barich, S. N., Memon, I. P., & Siddiqi, R. A. (2022). Exploring ESL college students' perceptions about English language tests: A qualitative study. *Journal of Positive School Psychology*, 6(10), 1836–1846.
- [7] Boraie, D. (2023). Learners' perceptions of language proficiency, language test-taking strategies and emotional regulation in a test-taking context: A case study in an Egyptian EFL context. In *Perspectives on Language Assessment Literacy* (pp. 45–68). Routledge.
- [8] Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Longman.
- [9] Chen, L., & Al-Mansoori, H. (2025). Neurophysiological correlates of structural uncertainty: An ERP study of morphosyntactic processing under stress. *Journal of Neurolinguistics*, 73, 101–118.
- [10] Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62.

- [11] Daud, S. (2023). Students' and teachers' perceptions of diagnostic tests in English language learning. *JETL (Journal of Education, Teaching and Learning)*, 8(1), 45–52.
- [12] Desheng, C., & Varghese, A. (2013). Testing and evaluation of language skills. *Journal of English as a Second Language*, 1-10.
- [13] Dowens, M. G., Vergara, M., Tola, A. L., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870–1887.
- [14] Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia - Social and Behavioral Sciences*, 2(2), 3600–3604.
- [15] Fitriani, D. (2020). Assessment practices: Challenges and opportunities faced by EFL teachers. *Journal of English Education*, 9(1), 11–18.
- [16] Halim, N., & Ardiningtyas, S. Y. (2018). Difficulties faced by the students in answering TOEFL test questions. *Eternal (English, Teaching, Learning, and Research Journal)*, 4(2), 224–231.
- [17] Karimullah, I. W., & Mukminatien, N. (2022). Problems faced and strategies applied by test-takers in completing the TOEFL iBT test. *Journal of English for Academic and Specific Purposes*, 5(1), 105–123.
- [18] Mahmud, M. (2014). The EFL students' problems in answering the Test of English as a Foreign Language (TOEFL): A study in Indonesian context. *Theory and Practice in Language Studies*, 4(12), 2581–2587.
- [19] Meylinda, L. (2019). Exploring students' perception on difficulties of structure and written expression in TOEFL test: A case study. *JEELL*, 6(1), 1–10.
- [20] Sultana, N., & Amin, M. N. (2021). Students' perception on high-stakes English language testing in Bangladesh: Expectations vs reality. *Journal of Asia TEFL*, 18(2), 485–501.
- [21] Upshur, J. A. (1971). Objective evaluation of oral proficiency in the ESOL classroom. *TESOL Quarterly*, 5(1), 47-59.
- [22] Vargas, M. R., & Schmidt, J. K. (2025). The role of executive functions in morphosyntactic acquisition: Cognitive load and perceived difficulty in EFL testing. *Language Learning & Technology*, 29(1), 45–63.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **CJHES** and/or the editor(s). **CJHES** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.