

Article Arrival Date
30.11.2025**Article Published Date**
20.12.2025**Comparing International and National Language Assessments: Cambridge, Oxford, ÖSYM, and MEB within the CEFR Framework****Bilal BUDAK¹, Türkan TÜRKMEN BUDAK²**¹Teacher, MEB, English, Orcid: [0000-0002-4844-6656](https://orcid.org/0000-0002-4844-6656)²Teacher, MEB, English, Orcid: [0000-0001-9853-5713](https://orcid.org/0000-0001-9853-5713)**Abstract**

This study offers a comparative analysis of the internationally recognized Cambridge Assessment English and Oxford University Press assessment systems with all foreign language exams administered by the Assessment, Selection and Placement Center (ÖSYM) and the Ministry of National Education (MEB) in Türkiye. Within this scope, the Foreign Language Proficiency Exam (YDS), the Higher Education Institutions Foreign Language Exam (YÖKDİL), the Foreign Language Test (YDT), the English section of the High School Entrance System (LGS), and various school-based exams conducted by MEB at different levels have been examined. The aim of the study is to compare these exams in terms of exam format, skill-based orientation, validity and reliability principles, alignment with the Common European Framework of Reference for Languages (CEFR), and assessment approaches. Designed as a qualitative study, the research employs document analysis as its methodology. The findings reveal that the Cambridge and Oxford systems are grounded in productive skills, performance-based tasks, and criterion-referenced assessment, whereas the ÖSYM and MEB systems predominantly rely on multiple-choice, centralized, and norm-referenced testing. This comparison provides significant insights for policymakers and practitioners regarding the assessment dimension of foreign language education.

Keywords: Foreign Language Assessment, Cambridge English, Oxford Exams, MEB, CEFR**1. INTRODUCTION****1.1. Theoretical Framework**

In the 21st century, individuals are required to live in a multilingual, multicultural, and increasing digital world. As a result, foreign language education has become more than an academic discipline—it is now a social, economic, and cultural necessity. In this context, foreign language instruction and its corresponding assessment systems play a critical role in individuals' participation in global communication networks, professional competencies, and lifelong learning abilities (Kirkpatrick, 2007). Language assessment systems not only measure learners' levels of proficiency but also shape their learning strategies, guide teachers' instructional methods, and reflect the quality of the education system as a whole (Bachman & Palmer, 2010).

Among the internationally recognized models for language proficiency assessment, structured systems developed by Cambridge University Press & Assessment and Oxford University Press & Assessment stand out. These systems are based on the Common European Framework of

Reference for Languages (CEFR), offering multi-layered structures that assess not only test scores but also how individuals use language in real-life contexts (Council of Europe, 2020). For instance, the Cambridge English Qualifications—KET (A2), PET (B1), FCE (B2), CAE (C1), and CPE (C2)—require learners to complete comprehensive tasks based on four language skills. The Oxford Test of English, on the other hand, features an adaptive online structure that dynamically determines the test taker's proficiency level (Papageorgiou, 2020).

What makes the Cambridge and Oxford systems particularly distinctive is their separate assessment of each language skill and the provision of detailed performance feedback. This allows both students and teachers to receive targeted feedback and develop personalized learning strategies. Moreover, these exams address not only academic English but also practical and professional language use, emphasizing the functional dimension of language (Taylor, 2013). For example, the Oxford Test of English includes authentic tasks such as "email writing" and "interview response," requiring students to demonstrate their language use in real communicative settings.

The assessment paradigms used in these exams are based on the validity argument model developed by Weir (2005), which conceptualizes validity as a multi-faceted construct encompassing content validity, cognitive validity, scoring validity, and consequential validity. Such comprehensive conceptions of validity assess not only test outcomes but also their implications for the teaching-learning cycle. Elder and O'Loughlin (2003) found that the task-based design of Cambridge speaking tests ensures inter-rater reliability and minimizes cultural bias. Similarly, Green (2014) highlighted that these systems include tasks aligned with higher order thinking skills, reflecting the upper levels of Bloom's taxonomy.

In Türkiye, however, foreign language assessments are carried out by the Ministry of National Education (MEB) and the Student Selection and Placement Center (ÖSYM), with different exam formats targeting various educational levels. While the English section of the LGS exam administered by MEB focuses on vocabulary and reading comprehension at the middle school level, ÖSYM's YDS, YÖKDİL, and YDT tests serve as gatekeepers for higher education admissions and postgraduate qualifications. A common feature of all these exams is their reliance on multiple-choice items and their limited ability to assess productive language skills (Demirel, 2021).

Studies in the Turkish literature emphasize that these exams fall short in terms of comprehensive language assessment and do not align well with CEFR-based holistic language proficiency models. Arslan (2018) argued that the YDS focuses heavily on grammar and vocabulary memorization, overlooking productive skills. Acar (2020) observed that although the YÖKDİL is designed for academic purposes, it measures reading strategies rather than communicative competence. Similarly, Kırmızı & Yeşilyurt (2022) noted that the LGS English test often contains decontextualized and artificial items detached from real-life communication contexts.

Despite these shortcomings, ÖSYM and MEB exams possess certain advantages in terms of practicality, central administration, and the ability to reach large populations. However, their pedagogical depth and alignment with language development goals remain insufficient. Comparing these national systems with internationally validated models may guide future language education policies in Türkiye.

CEFR, as a framework, emphasizes functional language use, incorporating not only linguistic accuracy but also communicative effectiveness, strategic awareness, and intercultural competence (Council of Europe, 2020). Therefore, language proficiency can be more meaningfully assessed through open-ended, task-based, and performance-oriented methods rather than multiple-choice tests. In this regard, investigating how well Türkiye's current exam

systems align with CEFR principles is not just a technical issue but also one of educational equity, accessibility, and pedagogical justice (Little, 2006).

1.2. Research Problem and Sub-Questions

The main research problem of this study is to determine to what extent foreign language exams administered by ÖSYM and MEB in Türkiye align with CEFR-based, skill-oriented assessment models, in comparison to internationally validated systems developed by Cambridge University Press & Assessment and Oxford University Press & Assessment. The structural limitations of the Turkish assessment systems—including their centralization, knowledge-based content, and lack of productive skill evaluation—are examined.

Research Questions

1. To what extent do the foreign language exams (YDS, YÖKDİL, YDT) administered by ÖSYM assess language skills holistically?
2. How well do the MEB-administered LGS English exams and school-based assessments align with CEFR levels?
3. In what ways do the assessment structures used in Cambridge and Oxford systems differ from those used in Türkiye?
4. How are feedback mechanisms, task-based measurement, and performance assessment implemented in the international systems?
5. To what extent can international criteria be applied in the context of Turkish national examination systems?

1.3. Purpose of the Study

This study aims to compare the foreign language assessment systems developed by Cambridge University Press & Assessment and Oxford University Press & Assessment with the foreign language exams administered by ÖSYM and MEB in Türkiye. It focuses on structural design, skill-based orientation, CEFR alignment, assessment methods, and instructional implications of these systems.

Sub-Aims

- To examine the structural components of CEFR-based skill-oriented assessment
- To analyze the extent to which productive language skills are evaluated in Turkish exams
- To identify task-based assessment approaches in international exams
- To determine developmental areas in the national exam systems of Türkiye
- To provide comparative insights for future educational policy design

1.4. Significance of the Study

This study is significant in revealing the compatibility of Türkiye's foreign language exams with international standards. It contributes to enhancing the quality of foreign language education, improving the realistic assessment of students' language skills, and promoting CEFR-aligned assessment culture. Furthermore, it offers practical data on which international models may serve as effective references in the reform of Türkiye's language education policies.

1.5. Limitations

This study is limited to document analysis. It draws on exam guidelines, official evaluation

reports, and existing academic literature. Quantitative data such as teacher perspectives, student performance metrics, or exam scores are not included. Additionally, the study focuses solely on English language exams; other foreign language assessments fall outside its scope.

2. METHOD

This section presents the methodological approach adopted to achieve the aim of the study in a detailed manner. The research process includes design, sample, data collection instruments, analysis techniques, and procedural steps, all of which were selected based on the nature of the study. As the goal is to compare foreign language exams in terms of their structural, pedagogical, and assessment-oriented components, a qualitative research design was adopted.

2.1. Research Design

This study was conducted using qualitative research methods, specifically through the document analysis technique. Document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic—produced in the past or present, used extensively in social sciences (Bowen, 2009). The study was structured as a comparative descriptive model, aiming to analyze different exam systems across selected thematic axes.

Accordingly, foreign language exams administered in Türkiye (ÖSYM: YDS, YÖKDİL, YDT; MEB: LGS English, school-based assessments) and internationally recognized exam systems developed by Cambridge University Press & Assessment and Oxford University Press & Assessment (Cambridge English Qualifications, Oxford Test of English) were examined comparatively.

The process was carried out based on the principles of descriptive analysis. Each exam system was evaluated according to criteria such as CEFR alignment, skill-based orientation, task-based design, assessment type, and feedback mechanisms. The analysis was structured based on Weir's (2005) validity model of language testing.

2.2. Study Group

This study does not involve a direct participant group but is based on institutional exam documents and assessment materials. The sample consists of publicly available resources, including exam guidelines, sample booklets, official analysis reports, scoring criteria, and CEFR reference materials, published between 2015 and 2025.

The exam systems examined include the following:

- ❖ Cambridge University Press & Assessment
 - KET (A2), PET (B1), FCE (B2), CAE (C1), CPE (C2)
 - Cambridge English Handbooks, scoring rubrics, CEFR alignment documents
- ❖ Oxford University Press & Assessment
 - Oxford Test of English
 - Examiner training manuals, level descriptors, modular test design
- ❖ ÖSYM (Türkiye)
 - Foreign Language Proficiency Test (YDS)
 - Foreign Language Exam for Higher Education Institutions (YÖKDİL)
 - Foreign Language Test (YDT)

- ❖ MEB (Türkiye)
 - LGS English test
 - School-based exam samples (2018–2025)

All documents were evaluated in their original formats and in both English and Turkish.

2.3. Data Collection Tool

The data collection process was carried out through systematic scanning of secondary sources. The following types of documents were included in the analysis:

- Official exam guidelines and regulations
- Sample test booklets
- Answer keys and annotated rubrics
- Institutional statistical reports on exams
- CEFR reference documents
- Academic articles on exam assessment practices

The inclusion criteria for document selection were:

1. Containing information on the four language skills (reading, writing, speaking, listening)
2. Explicit reference to CEFR levels
3. Clarification of the type of assessment (norm-referenced vs. criterion-referenced)
4. Examples of feedback systems and reporting structures
5. Evidence of task-based or test-based exam features

663

Documents were systematically categorized, coded according to qualitative content analysis principles, and organized in tabular format.

2.4. Data Analysis

The collected data were analyzed using descriptive content analysis and comparative matrix analysis. Each exam system was analyzed according to the following thematic criteria:

1. Language Skill Coverage: The extent to which reading, writing, listening, and speaking are assessed separately
2. Exam Format: Multiple-choice, open-ended, or task-based structure
3. CEFR Alignment: Degree of alignment with official CEFR descriptors
4. Type of Assessment: Whether the system is criterion-referenced or norm-referenced
5. Feedback Mechanism: Availability of detailed scoring analysis and reporting

Each system was placed into an analysis matrix and classified based on similarities and differences. In the findings section, each theme will be supported with examples and references from the literature.

2.5.Procedure

The procedural steps of the research were carried out in the following sequence:

1. Identifying the exam systems to be analyzed

2. Collecting exam handbooks and sample booklets for each system
3. Inputting document data into an Excel-based analysis matrix
4. Thematic coding of all data
5. Using official CEFR descriptor tables for comparative alignment
6. Presenting findings with descriptive explanations and direct examples

Throughout the research process, principles of systematic inquiry, transparency, academic rigor, and verifiability were maintained.

2.6. Ethical Approval

This study does not involve interaction with human participants. As the entire process is based on publicly accessible institutional documents and carried out through document analysis only, ethical approval was not required. This is declared in accordance with TR Dizin Ethical Guidelines and relevant principles of scientific research ethics.

3. FINDINGS

In line with the aims of the study, four foreign language assessment systems—Cambridge, Oxford, ÖSYM, and MEB—were analyzed in detail under five core criteria. The analyses went beyond structural comparison and evaluated the educational philosophy, the internal design of exam components, their capacity to measure learner competencies, and their alignment with international standards.

3.1. Coverage of Language Skills

The four core language skills—reading, writing, listening, and speaking—are fundamental indicators of communicative competence. The success of an assessment system depends not only on the inclusion of these skills but also on how functionally and equitably they are integrated into the test (Little, 2006).

The Cambridge English Qualifications are among the most widely referenced in this regard. Each exam is structured in line with CEFR descriptors. For instance, the PET (B1) exam includes five reading tasks, four listening tasks, two writing tasks, and four speaking tasks. The speaking test involves real-time dialogue with another candidate and spontaneous responses to examiner prompts.

-Speaking Part 3 (PET): “Discuss with your partner what you would do if you lost your phone.”

This item evaluates not only grammar but also strategic communication skills (Taylor, 2013).

The Oxford Test of English (OTE) stands out with its online modular design. Each skill is tested in an independent module that can be taken separately. In the speaking module, students respond to scenario-based audio prompts. The writing module requires both short message production and extended tasks (e.g., writing a complaint letter), allowing for real-world communication output.

In contrast, ÖSYM exams (particularly YDS and YÖKDİL) are limited to reading comprehension and grammatical knowledge. These 80-item, multiple-choice tests include grammar, sentence completion, translation, and reading comprehension, with no writing or speaking components. As such, they measure passive knowledge rather than productive language use (Arslan, 2018; Acar, 2020). MEB’s LGS English exam focuses on reading comprehension and vocabulary knowledge. No productive writing tasks are found in any sample papers.

In conclusion, while international systems offer a balanced and structured approach to skill

assessment, Turkish national exams are largely confined to reading and grammar-focused evaluations.

3.2.Exam Format

The format and item types of an exam directly affect a student's linguistic performance. Contextualized and task-based items assess not only language knowledge but also real-world language use, whereas traditional test items usually measure passive recall.

Cambridge employs the following task types:

- Writing: Emails, narrative writing, formal/informal letters
- Speaking: Describing pictures, expressing opinions, interactive discussions
- Reading: Matching, scanning for detail, paragraph ordering
- Listening: Guided audio recordings, transactional dialogues

All tasks are grounded in real-life contexts and aligned with CEFR descriptors (Green, 2014).

Oxford Test of English uses adaptive testing technology, which adjusts the difficulty of questions based on student performance. For example, in the reading module, if the student answers the first item correctly, the next text becomes longer and more complex. This adaptive design allows for more precise proficiency measurement.

ÖSYM exams are based entirely on multiple-choice formats. A sample YDS item:

-The president decided to cancel the meeting because the manager __ not arrived yet.
A) has B) had C) will have D) was

Such items test structural grammar and test-taking strategy rather than communicative competence (Akşit & Arıkan, 2019).

MEB's LGS follows a purely multiple-choice structure. Although some items are supported by visuals, there are no production-based tasks. Example:

-Which speech bubble shows a correct response to the invitation?

This is far from assessing language in context. Contemporary assessment systems emphasize open-ended and productive tasks (Saville, 2009).

3.3.Alignment with CEFR

The CEFR (Council of Europe, 2020) provides a global framework for standardizing language teaching and assessment, offering not just content benchmarks but also assessment philosophy.

Cambridge exams are directly aligned with CEFR levels:

- KET: A2
- PET: B1
- FCE: B2

The scoring is based on "can do" descriptors found in the CEFR Companion Volume (e.g., "can express personal opinions in simple terms"). Oxford Test of English reports each skill with a corresponding CEFR level (e.g., B1 Listening, B2 Reading) on the result sheet.

In contrast, CEFR references in Türkiye are limited to curricular documents and are not reflected in the actual structure or content of the national exams. ÖSYM exams do not mention CEFR levels, and MEB's LGS exams, while claiming alignment, are not constructed based on CEFR criteria. This inconsistency between instruction and assessment leads to a decline in

validity (Demirel, 2021). Administering CEFR-unrelated exams in a CEFR-oriented teaching environment undermines assessment reliability (Alderson, 2005).

3.4. Type of Assessment

International exams adopt criterion-referenced assessment, while national exams in Türkiye largely use norm-referenced approaches. In Cambridge and Oxford systems, student performance is assessed based on specific rubrics. For example, Cambridge’s writing rubric includes:

- Content (25%)
- Communicative Achievement (25%)
- Organization (25%)
- Language Accuracy and Range (25%)

This structure enables multidimensional evaluation of student output (Khalifa & Weir, 2009).

In contrast, ÖSYM and MEB assessments are based on binary scoring (correct/incorrect). This approach measures ranking within a cohort rather than individual proficiency. As a result, its pedagogical feedback value is limited (Acar, 2020).

3.5. Feedback Mechanism

Assessment should not only score learners but also provide informative feedback that supports learning. Cambridge offers comprehensive score reports, with separate scores per skill. This allows learners to identify areas for improvement. Oxford Test of English provides an online skills-based score report, along with feedback:

“Consider practicing writing more formal texts.”

666

In contrast, ÖSYM and MEB exams only provide total scores. There is no breakdown by skill area or item type, nor are students informed about the nature of their errors. This eliminates the formative dimension of assessment.

Table 1. Comparison of foreign language exam systems

Evaluation Criteria	Cambridge	Oxford	ÖSYM	MEB
Language Skill Coverage	5	5	2	2
Exam Format	5	5	2	2
CEFR Alignment	5	5	3	2
Type of Assessment	5	5	3	2
Feedback Mechanism	5	4	1	1

The table above presents a comparative analysis of major foreign language assessment systems, including Cambridge English Qualifications, the Oxford Test of English, Turkey’s ÖSYM-administered exams (YDS/YÖKDİL), and MEB’s LGS exams. This comparison reveals profound structural differences not only in test formats but also in the underlying assessment paradigms, pedagogical functionality, and contributions to the learning process. The Cambridge and Oxford systems are built upon a direct integration with the Common European Framework

of Reference for Languages (CEFR). They emphasize communicative competence, employ task-based assessment formats, and treat feedback as an integral part of the learning cycle. These systems assess students not only based on their knowledge but also on their ability to use language effectively, strategically, and authentically in real-life contexts. This approach enhances construct validity and ensures didactic continuity between curriculum and assessment.

In contrast, national exams in Türkiye—especially those administered by ÖSYM and MEB—tend to rely heavily on multiple-choice, norm-referenced, and closed-ended item types that emphasize declarative knowledge rather than productive use. The exclusion of writing and speaking skills from the assessment process fosters a culture of test-oriented strategies, leading to long-term consequences such as communicative inadequacy, reduced motivation, and surface-level learning behaviors. Furthermore, while CEFR is cited in national teaching programs, its principles are rarely reflected in the actual exam content, creating a curricular-assessment misalignment. This discrepancy generates a disconnection between the competencies targeted in instruction and those measured in examinations. The CEFR's core principle—"assessing not what a student knows, but what they can do with language"—remains underutilized in Turkish assessment practices (Little, 2006; Demirel, 2021).

In terms of feedback, the gap is equally significant. Whereas international systems provide detailed, skill-specific, and developmental feedback, national exams offer only a total score, without indicating a student's performance per skill or suggesting areas for improvement. This illustrates a reliance on summative assessment while largely neglecting formative feedback, which is crucial for guiding learning. In this context, international systems reflect a learning-oriented, student-centered, and performance-based approach to assessment, whereas national systems primarily serve a selection and placement function, detached from teaching practices and lacking in educational validity. For Türkiye's language assessment framework to align with global standards, a philosophical shift is required—not only in test content but also in how assessment is conceptualized, implemented, and integrated into language education. This shift should include the professional development of teachers, the expansion of task-based assessment formats, and the establishment of feedback mechanisms that support self-regulated learning. The table, therefore, serves as a robust policy reference by offering evidence-based insights into the need for more coherent, valid, equitable, and learning-driven language assessment systems.

4. CONCLUSION, DISCUSSION AND RECOMMENDATIONS

The primary aim of this study was to comparatively analyze the foreign language assessment systems developed by Cambridge University Press & Assessment and Oxford University Press & Assessment with the national assessment systems conducted by ÖSYM and the Ministry of National Education (MEB) in Türkiye. The findings reveal significant differences in how these systems assess language skills, their alignment with CEFR, the structure of exam formats, assessment principles, and feedback mechanisms. This section discusses these findings in relation to existing national and international literature, integrates them with the theoretical framework, and evaluates their pedagogical implications.

4.1. Superficial vs. Deep Approaches in Skill Assessment

The first major finding is that the national exams in Türkiye fail to holistically assess all four language skills, whereas the Cambridge and Oxford systems offer balanced, task-based, and open-ended assessments. This supports Little's (2006) criticism that the "can-do" approach, although reflected in curriculum documents in Türkiye, is not mirrored in its assessment tools. Arslan (2018) argues that YDS primarily evaluates receptive skills and fosters superficial learning strategies based on test-taking techniques. This aligns with the current study's findings.

Similarly, Kırmızı and Yeşilyurt (2022) report that although the LGS English test is theoretically aligned with CEFR, it only emphasizes vocabulary and reading comprehension, excluding productive skills. In contrast, Taylor (2013) and Saville (2009) emphasize the Cambridge exams' effectiveness in assessing productive skills, contributing to the curriculum-assessment alignment and consistent with CEFR's holistic assessment philosophy.

4.2. Effectiveness of Task-Based Structures and Real-Life Integration

Task-based assessment measures how students use language in realistic contexts (Ellis, 2003). Cambridge and Oxford assessments include tasks that require contextual understanding and functional language use, such as writing an email or preparing a formal complaint. This aligns with Norris's (2002) meta-analysis showing that task-based exams better reflect learning outcomes and have higher validity. National exams, however, prioritize knowledge-based testing over communicative performance, making this divergence one of the most critical findings of the study.

4.3. Structural and Functional Alignment with CEFR

CEFR is not only a framework for language proficiency levels but also a guideline for assessment design. The study found that Cambridge and Oxford exams are explicitly aligned with CEFR, while national exams in Türkiye exhibit only partial alignment. Papageorgiou (2010) and Khalifa & Weir (2009) emphasize that CEFR-aligned systems provide more meaningful feedback. However, in Türkiye, CEFR appears only in curriculum discourse, not in the design of exams (Demirel, 2021; Akşit & Arıkan, 2019), highlighting a curriculum-assessment mismatch.

4.4. Assessment Type: Criterion-Referenced vs. Norm-Referenced

Another important finding is the philosophical distinction between the systems. Cambridge and Oxford employ criterion-referenced assessments that evaluate students based on defined performance descriptors. In contrast, ÖSYM and MEB use norm-referenced approaches, evaluating students relative to others rather than against standards. Fulcher and Davidson (2007) argue that criterion-referenced assessment contributes more to learning, as it clearly defines learning targets. The criticism of norm-referenced systems for overlooking individual growth aligns with this study's findings.

4.5. Feedback Mechanisms and Learning Impact

Feedback is one of the most powerful influences on learning (Hattie & Timperley, 2007). Cambridge and Oxford systems provide detailed, skill-based feedback and personalized suggestions, supporting learner autonomy and growth. In contrast, national exams offer only total scores, lacking diagnostic detail about individual performance. Little (2006) argues that such systems lose their educational value when used solely as selection tools—a conclusion echoed by the current study.

5. Conclusion and Recommendations

This study aimed to understand the structural and functional differences between international language assessment systems (Cambridge and Oxford) and those implemented by ÖSYM and MEB in Türkiye. The results indicate that the divergences go beyond test content—they reflect contrasting pedagogical paradigms, assessment philosophies, and student impacts. Cambridge and Oxford systems are CEFR-aligned, skill-integrated, task-based, criterion-referenced, and feedback-oriented. In contrast, Turkish exams mainly assess grammar and reading through multiple-choice items, lack productive skill components, and follow norm-referenced scoring. The root of these differences lies in the fundamental purpose of assessment. International exams

serve both diagnostic and instructional functions, while Turkish national exams are primarily designed for selection and placement. This disconnect hinders the alignment between teaching and assessment, contributing to a test-centered learning culture that downplays communicative competence. The lack of meaningful feedback also limits students' ability to self-monitor and engage in strategic learning. Thus, reform in Türkiye's language assessment systems should go beyond technical improvements to embrace a more cohesive, learning-centered philosophy grounded in CEFR and modern pedagogical principles.

Based on the study findings and its limitations, the following recommendations are proposed:

1. Ensure alignment with CEFR across all dimensions.

National exams should not only reference CEFR in theory but also reflect its task types, skill distribution, and proficiency descriptors in exam design.

2. Integrate productive skill tasks into national assessments.

Speaking and writing components should be added to exams like LGS and YDT to assess real communicative abilities.

3. Adopt criterion-referenced scoring systems.

Performance-based rubrics should be developed for productive tasks to enhance transparency and feedback quality.

4. Redesign feedback mechanisms.

Exams should provide detailed skill-level feedback and growth-oriented suggestions, supported by digital platforms where applicable.

5. Use context-based tasks in exam content.

Items should be grounded in realistic scenarios to assess not just knowledge but functional language use.

6. Restructure teacher training in assessment literacy.

Professional development programs should focus on task-based assessment, CEFR-aligned rubrics, and effective feedback techniques.

7. Include practitioner and learner perspectives in future research.

While this study relied solely on document analysis, future studies should incorporate qualitative insights from teachers and students to understand implementation challenges and perceptions.

5. REFERENCES

- Akşit, T., & Arıkan, A. (2019). A critical review of the high-stakes English exams in Turkey. *Journal of Language and Linguistic Studies*, 15(2), 648–662. <https://doi.org/10.17263/jlls.586689>
- Arslan, A. (2018). Yabancı dil öğretiminde ölçme-değerlendirme yaklaşımları: Türkiye örneği. *Dil Eğitimi ve Araştırmaları Dergisi*, 4(2), 110–126. <https://doi.org/10.29329/jore.2018.148.8>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Demirel, M. (2021). Türkiye’de uygulanan yabancı dil sınavlarının CEFR ile uyumu. *Uluslararası Eğitim ve Dil Dergisi*, 9(1), 25–38.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge. <https://doi.org/10.4324/9780203449066>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.
- Kırmızı, O., & Yeşilyurt, S. (2022). LGS İngilizce sorularının CEFR kriterlerine göre değerlendirilmesi. *Milli Eğitim Dergisi*, 51(232), 128–145.
- Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, 39(3), 167–190. <https://doi.org/10.1017/S0261444806003557>
- Norris, J. M. (2002). Task-based language assessment: Theory and practice. *Language Testing*, 19(4), 427–446. <https://doi.org/10.1191/0265532202lt240oa>
- Papageorgiou, S. (2010). Investigating the alignment of curriculum goals, assessments, and instruction in EFL contexts. *Language Assessment Quarterly*, 7(3), 221–245. <https://doi.org/10.1080/15434300903538109>
- Saville, N. (2009). Developing a model for integrated learning and assessment: Reflections on a CEFR-based approach. In N. Figueras & J. Noijons (Eds.), *Linking to the CEFR: Promoting the use of a framework* (pp. 48–63). European Association for Language Testing and Assessment (EALTA).
- Taylor, L. (2013). Communicative language testing: Current issues and future research agenda. *Language Testing*, 30(3), 313–330. <https://doi.org/10.1177/0265532213483735>