

Journal of Language Education and Research (JLERE)

Dil Eğitimi ve Araştırmaları Dergisi (DEvAD)

ISSN: 2149-5602 | e-ISSN: 2149-5602

2026 | 12 (1) | 867-886 |

DOI: 10.31464/jlere.1902199

Article Type: Research Article

Suggested Citation

Derer, Ö. K., Batmaz Derer, N., & Berkant, H. G. (2026). Middle school students' perceptions of listening and speaking performance exams in English: A qualitative case study at a Turkish state school. *Journal of Language Education and Research*, 12 (1), 867-886. <https://doi.org/10.31464/jlere.1902199>

Article History

Received: 06.03.2026

Revised: 18.08.2026

Accepted: 25.08.2026

Published: 20.09.2026

Author Contribution Statement

Author 1: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft Preparation.

Author 2: Data Curation, Investigation, Formal Analysis, Writing – Review & Editing, Validation.

Author 3: Supervision, Methodological Guidance, Review & Editing.

Ethical Statement

This study was approved by the Yozgat Bozok University Ethics Committee by the decision dated: June 28, 2024, and numbered [Decision No. 15/42].

Declaration of Generative AI Use

Generative artificial intelligence was utilized to improve language clarity and correct grammatical and stylistic issues in the manuscript.

Middle School Students' Perceptions of Listening and Speaking Performance Exams in English: A Qualitative Case Study at a Turkish State School

Özgür Kürşad Derer^{1*}, Nuriye Batmaz Derer², Hasan Güner Berkant³

¹ Ministry of National Education, Meliha Ercan Secondary School, Bilecik, Türkiye — ORCID: 0000-0002-2383-5540

² Ministry of National Education, Bozüyük Science and Art Center, Bilecik, Türkiye — ORCID: 0000-0003-4279-9159

³ Yozgat Bozok University, Faculty of Education, Department of Educational Sciences, Curriculum and Instruction, Yozgat, Türkiye — ORCID: 0000-0003-0725-6036

Corresponding author's email: ok.derer@outlook.com.tr

Abstract

Recent curriculum reforms in Türkiye have emphasized communicative competence and performance-based assessment in English as a Foreign Language (EFL) yet listening and speaking assessments remain challenging to implement in exam-oriented school contexts. This qualitative case study explores middle school students' perceptions of listening and speaking performance exams in English lessons, focusing on perceived fairness, emotional responses, exam design, and perceived effects on learning and classroom participation. Data were collected through semi-structured, face-to-face interviews with 20 students (grades 5–8) attending a public middle school in Bozüyük, Bilecik, Türkiye. Interviews lasted 15–25 minutes, were audio-recorded with consent, and transcribed verbatim. Data were analyzed using content analysis with iterative coding and peer debriefing, and intercoder agreement was high (0.93). Findings reveal a central tension. Many students viewed performance exams as more authentic than paper-based tests and reported increased participation and greater awareness of listening and speaking development. However, the same exams were experienced as emotionally demanding, often associated with anxiety, time pressure, uncertainty about expectations, and concerns about grading fairness. Students reported distinct difficulty profiles across skills, including speech rate, unfamiliar vocabulary, and accent variation in listening, and anxiety-related hesitation, sentence construction under pressure, and word retrieval difficulties in speaking. Facilitating factors included clear instructions, opportunities for repetition, topic familiarity, and supportive classroom practice. Overall, the results suggest that perceived fairness and effectiveness depend on instructional alignment, transparent criteria, predictable procedures, and anxiety-sensitive conditions. The study highlights the value of student voice for strengthening assessment design and informing teacher training in Turkish EFL settings.

Keywords: *EFL middle school learners, listening assessment, performance-based assessment, speaking assessment, student perceptions*

1. Introduction

In the most recent national high school English language curriculum, communication remains the primary goal of language learning; however, the curriculum has been criticized for failing to adequately implement an action-oriented perspective that would position learners as social actors engaged in meaningful social action (Acar, 2025). This orientation reflects broader trends in English as a Foreign Language (EFL) pedagogy that emphasize action-oriented competence and learners' social agency (Council of Europe, 2001, 2020), moving away from grammar-focused models toward an approach in which learners use language to accomplish meaningful real-world tasks. Within this view, the balanced development of the four language skills, listening, speaking, reading, and writing, is regarded as essential for effective language acquisition (Council of Europe, 2001). In the Turkish context, concerns have also been raised about teachers' skill-specific assessment knowledge, particularly in listening, while assessing speaking has likewise been identified as a challenging area (Ölmezer-Öztürk & Aydın, 2019).

In response, the Ministry of National Education (MoNE) has sought to align assessment practices with the Common European Framework of Reference for Languages (CEFR) (Fındıklı & Büyükkarcı, 2024). A key element of this reform is the integration of listening and speaking performance exams intended to mirror authentic communication and to evaluate productive and receptive skills more holistically than multiple-choice formats allow (Council of Europe, 2001; MoNE, 2023).

Middle school learners are a strategically important population within this landscape (Özcanlı & Kozikoğlu, 2024). At this stage, peer evaluation and fear of making mistakes can become important sources of foreign language anxiety (Köksoy & Deniz, 2025). Learners at this age benefit from contextualized environments that integrate language acquisition with meaningful interaction. Task-based approaches have also been associated with stronger motivation and learning outcomes (Ellis, 2003). Listening and speaking are central to this orientation because they underpin communicative competence (Richards, 2006). Speaking, in particular, is closely associated with learners' confidence and motivation (Öztürk & Gürbüz, 2014).

From the learners' perspective, speaking exams can provoke high levels of anxiety, especially when students perform in front of peers or respond on the spot (Horwitz et al., 1986). Öztürk and Gürbüz (2014) report that although Turkish EFL learners may indicate a relatively low overall level of speaking anxiety, they perceive tasks involving immediate questions and lack of preparation as highly anxiety-provoking, sometimes accompanied by physiological reactions such as increased heart rate. In the Turkish context, traditional testing and assessment practices may remain misaligned with the newer learning cultures emphasized in curriculum reforms (Hatipoğlu, 2015). Students' experiences at this stage are also shaped by contextual factors such as peer attitudes, teacher behaviour, and the broader learning environment, which can either alleviate or intensify foreign language anxiety (Köksoy & Deniz, 2025; Öztürk & Gürbüz, 2014).

Studies have documented variability in English teachers' skill-specific assessment knowledge, with lower levels particularly in listening and writing assessment (Satgun & Aşık, 2023), have compared EFL teachers' assessment knowledge across contexts in relation to communicative goals (Fındıklı & Büyükkarcı, 2024), have examined English speaking self-efficacy perceptions (Özcanlı & Kozikoğlu, 2024) and have analyzed the cognitive, metacognitive, and socio-affective strategies that shape listening comprehension (Gilakjani & Sabouri, 2016). Far fewer studies have examined how middle school students themselves interpret listening and speaking performance exams in real classroom settings, and student voice in assessment design and evaluation remains largely absent in the Turkish EFL context (Fındıklı & Büyükkarcı, 2024). This scarcity of learner-centered evidence leaves a gap between the communicative, performance-based goals stated in national curriculum documents and the realities of students' assessment experiences.

Because students are the primary stakeholders in assessment, their views on the design, implementation, and perceived impact of these exams are central to meaningful evaluation. The present study therefore investigates middle school students' perceptions of listening and speaking performance exams in a Turkish EFL setting, focusing on how students evaluate the fairness, difficulty, feedback, and emotional dimensions of these assessments. An in-depth qualitative approach within a single instructional context is adopted to capture these situated perceptions closely, positioning learners not merely as test takers but as participants whose perspectives can inform fairer and more pedagogically sound assessment. Accordingly, the study addresses the following problem: What are the perceptions of middle school students about listening and speaking performance exams in English lessons?

1.1. Conceptual Framework

This study is grounded in four interrelated constructs that frame how listening and speaking performance exams are designed, experienced, and evaluated in EFL settings: performance-based assessment, foreign language anxiety, language assessment literacy, and student voice. Together they move from the design rationale of such exams toward the learner-centered concern that motivates the present study.

Performance-based assessment refers to the evaluation of language ability through tasks that require learners to produce or process language in communicative situations rather than to select answers in discrete-point formats (Council of Europe, 2001). Within action-oriented frameworks, such assessment is intended to mirror authentic communication and to capture productive and receptive skills more holistically (Council of Europe, 2020). Listening and speaking are central to this orientation because they underpin communicative competence (Richards, 2006). Speaking, in particular, is closely associated with learners' confidence and motivation (Öztürk & Gürbüz, 2014).

Closely tied to oral performance is foreign language anxiety, a situation-specific form of anxiety that arises in language learning and use, particularly under conditions of evaluation (Horwitz et al., 1986). Speaking in a foreign language can be particularly anxiety-provoking because it involves communication apprehension and fear of negative evaluation (Horwitz et al., 1986). In the Turkish EFL context, Öztürk and Gürbüz (2014) found that learners reporting relatively low overall speaking anxiety nonetheless experienced tasks involving immediate questions and lack of preparation as highly anxiety-provoking, at times with physiological reactions. Such anxiety can be heightened by contextual factors including negative peer evaluation, fear of making mistakes, harsh teacher attitudes, and an unsupportive learning environment (Köksoy & Deniz, 2025; Öztürk & Gürbüz, 2014).

Whether such exams are administered coherently and fairly depends in turn on language assessment literacy, which concerns teachers' knowledge and skills in designing and implementing appropriate assessment. Research indicates that EFL teachers' skill-specific assessment knowledge varies across language skills, with particularly limited knowledge in listening assessment (Satgun & Aşık, 2023), and comparative work shows variability in teachers' assessment knowledge and in how closely their practices align with communicative goals, with implications for the fairness and validity of performance exams (Fındıklı & Büyükkarcı, 2024). These gaps may sustain a misalignment between the learning cultures promoted by curriculum reforms and traditional classroom assessment practices (Hatipoğlu, 2015).

These constructs converge on student voice, the inclusion of learners' perspectives in decisions about the design, implementation, and interpretation of assessment. In the Turkish EFL context, Hatipoğlu (2015) emphasized the importance of systematically eliciting learners' views and perceived needs when developing assessment-related training. Much of the existing Turkish EFL assessment literature has focused on teachers' assessment knowledge, needs, and practices (Ölmezer-Öztürk & Aydın, 2019), while far fewer studies have examined how students themselves interpret and experience listening and speaking performance exams. Foregrounding student voice is therefore essential for understanding whether these assessments are perceived as fair and meaningful, and for refining evaluation practices in line with the communicative goals they are meant to serve.

1.2. Purpose of the Study

The main aim of this study is to examine middle school students' perceptions of listening and speaking performance exams in English lessons. Specifically, the study explores how students perceive the effects of these exams on their English learning, how they perceive the design and evaluation of the exams, and how they perceive the influence of these exams on their engagement and affective responses in the classroom. In line with this aim, the study seeks to answer the following research questions:

1. How do middle school students perceive the effects of listening and speaking performance exams on their English learning, including their preparation and self-regulation processes?
2. How do students perceive the implementation, design features, and evaluation procedures of these exams?
3. How do students perceive the influence of these exams on their classroom participation, engagement, and affective responses?

2. Methodology

This section presents the research design, context, participant selection, data collection instruments and procedures, data analysis, and the measures taken to ensure trustworthiness and ethical integrity. The methodological choices were guided by the aim of obtaining rich, context-grounded insight into how middle school students perceive listening and speaking performance exams in English lessons. Given the exploratory and interpretive nature of the research questions, a qualitative orientation was adopted.

2.1. Research Design

This study adopted a qualitative case study design to explore students' perceptions of listening and speaking performance exams. Qualitative inquiry is well suited to understanding how participants interpret educational phenomena within their natural settings rather than quantifying predetermined variables (Merriam & Tisdell, 2016). A case study design was selected because the phenomenon under investigation is embedded within a specific educational context and cannot be meaningfully separated from its real-life conditions, allowing for detailed, holistic description of a bounded system (Creswell & Poth, 2018). Listening and speaking performance exams represent such a phenomenon, shaped by curricular policies, teacher practices, classroom dynamics, and students' cognitive and affective responses. In the present study, the case was defined as middle school students' perceptions of performance-based listening and speaking assessments within one public school, bounded by one academic year, one school setting, and students who had direct experience with both types of exams. This boundedness allowed for an intensive and context-sensitive exploration.

The study was structured as a holistic single-case design, in which one unit of analysis is examined comprehensively rather than divided into subcases (Tracy, 2020). Accordingly, the research sought to capture students' preparation, affective responses, perceptions of fairness, and views on exam design and outcomes as interconnected aspects of the same assessment experience. Semi-structured interviews served as the primary data collection method, offering consistency across participants while allowing the researcher to probe for clarification and elaboration (Creswell & Poth, 2018).

2.2. Research Context

The study was conducted in a public middle school in Bozüyük, Bilecik, Türkiye. In this setting, English is a compulsory foreign language taught from the early grades through regular weekly instruction.

In line with recent curricular reforms emphasizing communicative competence and performance-based assessment, traditional paper-based examinations are complemented by oral and listening performance tasks intended to evaluate students' ability to use English in communicative situations. Students therefore participate in both speaking and listening performance exams alongside conventional written tests.

Within this school, these performance-based assessments are a routine component of evaluation rather than occasional or experimental practices. Listening and speaking exams are implemented systematically throughout the academic year and contribute to students' grades. Students thus regularly comprehend spoken input, respond orally, and demonstrate communicative competence under assessment conditions. This consistent exposure makes the school an information-rich environment for examining how learners perceive such assessments. Examining students' perceptions within their natural learning environment is consistent with naturalistic inquiry, in which data collected in authentic settings are more likely to reflect participants' real experiences (Lincoln & Guba, 1985; Patton, 2015). Rather than eliciting hypothetical opinions about imagined scenarios, the study captured reflections grounded in students' ongoing classroom experiences, enabling them to refer to concrete events and recent exams and thereby producing more detailed and credible accounts.

2.3. Participants

Participants were 20 middle school students from grades 5 through 8 enrolled in the selected public school. This sample size was considered adequate for an in-depth qualitative inquiry, as the aim was not statistical generalization but rich, experience-based understanding. In qualitative case study research, smaller samples allow researchers to explore participants' perspectives comprehensively rather than producing broad but superficial findings (Merriam & Tisdell, 2016). Purposive sampling was used to select information-rich participants with direct experience of the phenomenon, as the goal was depth of understanding rather than representativeness (Patton, 2015).

Two complementary purposive strategies were employed. First, stratified purposive sampling ensured representation from each grade level, so that perceptions could be examined across grades 5 to 8 and in relation to variation associated with age and accumulated school and exam experience. Older students, for example, may have encountered more performance exams and developed more established coping strategies, whereas younger students may respond more uncertainly to assessment situations. Second, criterion sampling ensured that all participants met a key eligibility requirement, namely prior experience with at least one listening and one speaking performance exam in English, so that their reflections were grounded in concrete experience rather than hypothetical expectation (Palinkas et al., 2015). The participants represented diversity in gender (10 girls and 10 boys), English achievement scores (55 to 92 out of 100), and self-reported speaking confidence, as reflected in their accounts of comfort, hesitation, and anxiety during the interviews. This range of academic and affective profiles enriched the dataset, and the combined use of stratified purposive and criterion sampling produced an information-rich and contextually appropriate participant group.

2.4. Data Collection Instruments

Data were collected through a semi-structured interview form developed by the researchers. The guide consisted of eight open-ended questions aligned with the research questions and the conceptual framework, addressing students' preparation processes, implementation experiences, affective reactions, perceptions of fairness, feedback and evaluation practices, and reflections on learning outcomes. This multidimensional structure ensured that the interviews captured both cognitive and affective aspects of

students' assessment experiences. Semi-structured interviews were selected because they balance structure with flexibility: all participants respond to the same core questions, while the researcher can ask follow-up and clarification questions to explore emerging points in depth (Creswell & Poth, 2018). This conversational format also allows participants to express their views in depth while enabling the interviewer to maintain focus on the topics relevant to the study (Jamshed, 2014).

The interview form was developed through a systematic process. An initial pool of questions was generated from the research questions and relevant literature on performance-based assessment, and these were organized into thematic clusters corresponding to preparation, implementation, evaluation, and perceived impact, ensuring that each research question was represented. To enhance content and language validity, the draft guide was reviewed by two experts in English language education and qualitative research methods, who evaluated the items for clarity, age appropriateness, linguistic simplicity, and conceptual alignment. Based on their feedback, wording was simplified to match students' developmental levels, abstract terminology was reduced, and double-barrelled questions were separated into single-focus items. A pilot interview was then conducted with one student who met the sampling criteria but was not included in the main study, in order to evaluate the comprehensibility, sequencing, and flow of the questions. The pilot indicated that some items required minor rephrasing and reordering for a more natural progression, and these revisions were implemented before the main data collection. Sample items include "What do you think about having separate listening and speaking performance exams in your English class?" and "How do these listening and speaking exams affect your participation and success in English lessons?", with additional probes inviting students to elaborate on the difficulties they experienced and the strategies they used to prepare.

2.5. Data Collection Procedure

Data were gathered through individual, face-to-face interviews conducted on the school premises, allowing students to participate in a familiar setting without disrupting their routines. Interviews took place in a quiet counselling room to ensure privacy and minimize distractions. A calm and supportive environment was considered particularly important given the participants' age and the potentially sensitive nature of discussing exam-related anxiety, stress, or perceived unfairness, as a private setting reduced peer influence and enabled students to express themselves more freely (Merriam & Tisdell, 2016). Notably, the researcher is a teacher at the same school but did not teach any of the participating students, which supported familiarity with the school context while minimizing the power dynamics that can arise when an interviewer is also the students' own evaluator. Before each interview, brief rapport-building conversations were held, and the researcher explained that there were no right or wrong answers and that the aim was to understand students' experiences rather than to evaluate them.

Each interview lasted between 15 and 25 minutes, depending on the depth of students' responses. The semi-structured format allowed some interviews to extend when participants provided detailed narratives or when probing was needed. Prior to the interviews, participants were informed about the purpose of the study, the voluntary nature of participation, and their right to withdraw at any time without consequences, and they were assured that their responses would remain confidential and would not affect their grades or their relationships with teachers. These explanations were provided in clear, age-appropriate language, and written parental consent and student assent were obtained. With permission, all interviews were audio-recorded and later transcribed verbatim, preserving participants' original wording. The transcription process yielded approximately 70 pages of raw textual data, providing a substantial corpus for in-depth analysis and enabling repeated readings and systematic coding.

2.6. Data Analysis

The interview data were analyzed using qualitative content analysis, which enables researchers to systematically code textual data, examine meaning, and develop interpretations grounded in participants' perspectives (Vaismoradi et al., 2016). The analysis combined inductive content analysis with thematic analysis: content analysis guided the systematic segmentation and coding of the data, while the principles of thematic analysis informed the construction and review of themes (Braun & Clarke, 2006). This approach was appropriate because the aim was to explore students' perceptions and to develop themes inductively from the data.

The analysis followed a multi-stage and iterative process. First, all interviews were transcribed verbatim and read several times to achieve immersion, with preliminary notes taken on recurring ideas and notable statements. Second, meaningful units, consisting of words, phrases, or sentences conveying a distinct idea related to the research questions, were identified and assigned initial codes derived directly from the data rather than from a predefined framework. Third, codes sharing conceptual similarities were grouped into broader categories, with constant comparison employed across transcripts and codes refined, merged, or separated to enhance conceptual clarity. Fourth, broader themes were constructed by synthesizing related categories into higher-level interpretations that captured how students perceived listening and speaking performance exams, with each theme reviewed and revised until it demonstrated internal coherence and clear distinction from others (Braun & Clarke, 2006).

The process was recursive rather than linear, as the researchers moved back and forth between transcripts, codes, categories, and themes to verify interpretations. To enhance transparency and reflexivity, analytic memos documenting coding decisions and emerging insights were maintained throughout, creating an audit trail. In addition, a second researcher independently coded the data to evaluate the consistency of the coding scheme, and discrepancies were discussed until consensus was reached, strengthening the credibility and dependability of the findings.

2.7. Trustworthiness

Trustworthiness was established in accordance with Lincoln and Guba's (1985) framework, which addresses rigor through credibility, transferability, dependability, and confirmability.

Credibility was enhanced through several strategies (Creswell & Poth, 2018; Lincoln & Guba, 1985; Shenton, 2004). First, prolonged familiarity with the research setting was supported by the researcher's position as a teacher working at the school throughout the academic year, which fostered familiarity with the school context and the assessment practices under study, even though the researcher did not teach the participating students and data were collected exclusively through interviews. Second, peer debriefing was conducted throughout the analysis, as a second researcher reviewed transcripts, coding decisions, and thematic interpretations and provided critical feedback that helped challenge potential bias. Third, direct quotations from participants were included in the findings to preserve students' original voices and to demonstrate that interpretations were grounded in the data.

Transferability was addressed through thick description. Detailed accounts of the research setting, participant characteristics, sampling procedures, data collection, and analytic steps allow readers to judge the relevance of the findings to their own contexts (Creswell & Poth, 2018; Lincoln & Guba, 1985). Given that the study is based on 20 students in a single public school, the findings are offered as context-sensitive

insight rather than as evidence generalizable to public middle schools more broadly, and readers are positioned to assess their applicability to comparable settings.

Dependability was supported through systematic documentation of all methodological procedures, including instrument development, data collection, coding, and theme generation. Intercoder agreement was also calculated: two researchers independently coded the entire dataset of 200 meaningful units, and reliability was computed using Miles and Huberman's (1994) formula. Based on 186 agreements and 14 disagreements, the resulting coefficient was 0.93, indicating a high level of coding consistency.

Confirmability was ensured through a systematic audit trail. Analytic memos, coding records, decision logs, and reflective notes were preserved throughout the study, creating a transparent record of how interpretations developed and enabling the findings to be traced back to the original data (Carcary, 2009; Creswell & Miller, 2000; Nowell et al., 2017).

2.8. Ethical Principles

Ethical approval was obtained from the Ethics Committee of Bozok University (Decision No. 15/42, June 28, 2024) prior to data collection. All procedures complied with institutional ethical guidelines and national regulations concerning research involving human participants, particularly minors. Given that the participants were middle school students, special attention was paid to informed and voluntary participation. Written parental consent and student assent were obtained, and the purpose, procedures, and participants' rights were explained in clear, age-appropriate language. Students were explicitly informed that participation was voluntary and that they could decline any question or withdraw at any time without negative consequences.

To protect privacy and confidentiality, pseudonyms were assigned to all students, and no identifying information was included in transcripts, reports, or publications. Audio recordings, transcripts, and related documents were stored securely on password-protected devices accessible only to the researchers, and all data were used solely for academic purposes. Because discussing exam experiences might evoke stress or anxiety, a supportive and non-judgmental atmosphere was maintained, and students were reminded that the interviews were not evaluations and that there were no right or wrong answers.

2.9. Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

3. Findings

This section presents the findings of the qualitative analysis, organized around the study's three research questions. Drawing on semi-structured interviews, the findings reflect how middle school students perceive the effects of listening and speaking performance exams on their learning, how they perceive the design and evaluation of these exams, and how they perceive their influence on classroom participation and affective responses. Throughout, listening and speaking are presented as distinct skills, since students' accounts indicate that the two are associated with different demands and concerns.

3.1. Perceived Effects of the Exams on English Learning, Preparation, and Self-Regulation

This first theme addresses students' perceptions of how listening and speaking performance exams affect their English learning, together with the preparation and self-regulation processes through which they

respond to these exams. The findings show that students perceived the exams as consequential for their learning in both supportive and critical terms, and that they engaged in a wide range of preparation and improvement strategies that differed across the two skills.

3.1.1. Perceived Effects on English Learning

Students' perceptions of how the exams affected their learning clustered around the codes of perceived authenticity, limitations in assessment, contribution to improvement, impact on grades, and feelings of insecurity. A number of students described listening and speaking exams as more authentic and meaningful than traditional paper-based tests, perceiving performance tasks as closer to real-life language use and therefore better indicators of communicative ability. Ela explained that such tasks *feel more real than paper tests* and offer stronger opportunities to demonstrate communication skills. Derin stated that the exams helped her understand how English functions *in actual conversation, not just in exercises*, while Asya emphasized that these assessments made English feel *useful and alive*.

In contrast, several students highlighted limitations in assessment, particularly regarding fairness and accuracy. Emir noted that he sometimes performed poorly *not because I didn't understand, but because I was stressed*. Hasan referred to external variables such as background noise or unclear instructions, and Duru reported that *timing and pressure* negatively affected her performance. These accounts indicate that contextual and emotional factors were perceived to influence outcomes beyond students' actual language competence.

Other participants viewed the exams as supporting their development, coded as contribution to improvement. Mert stated that preparing for these tasks pushed him *to improve fluency and vocabulary in a practical way*. Kerem explained that the speaking components encouraged him *to practice more and reflect on mistakes*, while Efe emphasized that the exams *show progress over time*, especially when similar tasks were repeated. The code impact on grades also emerged strongly, as several students perceived performance tasks as high-stakes events. Can expressed that even minor mistakes could *drag your grade down a lot*, Melis noted that speaking days had *a big effect on the final score*, and Baran described these exams as *riskier than other parts*. Finally, many students expressed feelings of insecurity. Alp stated that he sometimes *knew what to say but froze when the exam started*, while İrem explained that *the pressure made everything feel harder* despite preparation. Taken together, these perceptions indicate that the exams were experienced not merely as assessment tools but as learning experiences shaped by pedagogical, contextual, and affective dimensions.

3.1.2. Preparation and Improvement Strategies for Listening

Students' preparation for the listening exam ranged from systematic study to lack of preparation and anxiety-driven behaviour, clustering around the codes of planned preparation, lack of preparation, practice activities, repeated listening, and anxious preparation. Several students described planned preparation. Ela explained that she created a checklist of topics and practiced listening tasks daily, while Melis listened to *old audio materials from class to refresh vocabulary and improve focus*. In contrast, some reported lack of preparation. Emir shared, *It's just listening, I thought I could handle it without practice*, while Alp explained that he simply forgot about the exam, saying *there were too many other things going on*. Others engaged in practice activities, such as completing online listening quizzes (Kerem) or, as Baran described, practicing with a friend by playing short clips and asking each other questions. Repeated listening was also frequent, as Can reported replaying classroom recordings to *train his ears* and Duru stated that repetition helped her notice missed details. Finally, several students described anxious preparation. Arda explained that he

panicked and tried everything at once the night before, and Naz admitted, I kept thinking I wasn't ready no matter how much I tried.

Beyond exam preparation, students described ongoing efforts to improve their listening, coded as repeated listening, video viewing, self-study, planned study, seeking support, and increased awareness. Ela noted, *When I listen to the same material more than once, I notice the words I missed the first time*, and Derin observed, *Repetition makes me more confident because I understand native speech better each time*. Under video viewing, Mert commented, *I regularly follow YouTube channels in English to improve both my listening and vocabulary*, and Asya said, *Short video clips help me stay focused and understand the language in context*. Self-study practices included Hasan's use of *mobile apps with listening exercises that match my level* and Kerem's self-made *routine using podcasts and songs*. Under planned study, Efe noted, *I organize both extensive and intensive listening tasks in my schedule*. Seeking support featured collaboration and guidance, as Zeynep explained, *I ask my teacher for extra materials and guidance*, and Duru stated, *Working with a classmate helps me understand difficult sections better*. Finally, increased awareness captured students' growing recognition of their weaknesses, as Can observed, *I realized I often miss key details, so now I pay closer attention in class*, and İrem reflected, *Understanding my limitations makes me try new strategies*. These accounts indicate that listening difficulties were addressed through comprehension-focused strategies, consistent with the receptive nature of the skill.

3.1.3. Preparation and Improvement Strategies for Speaking

Preparation for the speaking exam also varied widely, clustering around the codes of individual preparation, lack of preparation, practice, memorization, and anxious preparation. Several students reported individual preparation involving rehearsal and self-monitoring. Ela wrote short paragraphs at home and practiced delivering them aloud, Mert described using a mirror to improve fluency and confidence, and Asya prepared flashcards to support spontaneous speech. In contrast, some demonstrated lack of preparation, as Emir went into the exam without a specific plan and Hasan believed that regular speaking practice in class was sufficient. Others engaged in practice through peer-supported or self-monitored activities, such as Duru, who recorded her own responses and listened back to identify areas for improvement. The code memorization emerged often as a coping mechanism for uncertainty, as Kerem prepared by memorizing set phrases usable in multiple contexts and Naz used memorization to manage stress. Finally, several students described anxious preparation, as Alp prepared in a disorganized and rushed manner and İrem reported that she kept questioning her readiness.

In their ongoing efforts to improve speaking, students described strategies coded as practice, speaking in front of a mirror, memorization, vocabulary learning, English speaking videos, and self-awareness. Under practice, Ela stated, *I try to speak English as much as possible, even outside of school*, and Mert noted, *Practicing regularly reduces my hesitation and makes speaking feel more natural*. Mirror practice allowed students to monitor performance visually, as Emir shared, *When I speak to my reflection, I can notice my pronunciation and facial expressions*, and Melis added, *I can correct myself immediately while I'm speaking*. Under memorization, Arda explained, *I memorize key sentence patterns, so I feel more prepared*, and Kerem commented, *Memorizing useful expressions gives me a safety net when I get stuck*. Vocabulary learning supported oral production, as Hasan remarked, *Learning new words helps me to form my answers more confidently*, and Can noted, *Having a larger vocabulary gives me more flexibility when I speak spontaneously*. Under English speaking videos, Duru explained, *I watch interviews and dialogues to observe how people speak naturally*, and Asya added, *Imitating native speakers makes me feel more fluent*. Finally, self-awareness captured reflective efforts, as Defne stated, *I became more aware of my weak points and*

started tracking my progress, and İrem shared, *Realizing that I need to improve motivates me to keep working*. In contrast to listening, these strategies foreground production, rehearsal, and the management of speaking under evaluation, reflecting the productive and socially exposed nature of the skill.

3.2. Perceptions of the Implementation, Design, and Evaluation of the Exams

This second theme addresses how students perceived the way the exams were implemented, designed, and evaluated, including the separation of the two skills, the challenges and facilitators specific to each skill, the adequacy of exam duration and question load, and the fairness of the scores they received.

3.2.1. Perceptions of Assessing Listening and Speaking Separately

Students' views on assessing the two skills separately divided into positive and negative perceptions, with positive views coded as meaningful, supportive, easy, and facilitating focus, and negative views coded as unnecessary, difficult, and anxiety-inducing. Among those with positive perceptions, Ela found the separation meaningful, stating, *It makes sense to separate them because they are different skills and need different types of thinking*, and İlayda shared, *It helps me understand which part I'm good at*. Derin described it as supportive: *When they are separate, I can focus better and feel less pressure*, while Melis added, *It helps me concentrate more on one thing at a time*. Among those with negative perceptions, Emir found it unnecessary, stating, *We don't need two exams, one is enough to see if we can use English*, and Can found the format difficult, saying, *It's hard to switch between skills and remember everything*. Anxiety also recurred, as Ela, despite seeing the value, confessed, *Even though it makes sense, it still stresses me out*, and Duru said, *It makes me nervous because I don't know what to expect in each part*. These responses suggest that skill-specific exams were perceived to offer pedagogical benefits while also imposing emotional and cognitive demands.

3.2.2. Challenges and Facilitators in the Listening Exam

Students' experiences of the listening exam clustered into challenges, coded as meaning difficulty, speed difficulty, and accent difficulty, and facilitators, coded as repetition advantage, instruction clarity, and concentration. Under challenges, students reported difficulty in understanding the texts, as Ela struggled to catch the meaning when unfamiliar words appeared and Emir often got lost after missing a single sentence. Speed of delivery was also reported, as Derin stated the audio moved too fast to process and Melis added that fast speakers made her feel anxious and distracted. Some identified accent as a barrier, as Mert explained that different accents made it harder to recognize familiar vocabulary and Duru commented that she needed more exposure to different accents. Among facilitators, the repetition advantage was prominent, as Can noted that hearing the recording more than once allowed him to double-check what he heard and Kerem stated the second listening helped him notice missed details. Clarity of instructions also helped, as Asya explained that clear instructions gave her a sense of control. Finally, several students emphasized focus and concentration, as Naz prepared herself mentally before the listening began and Yiğit remarked that maintaining concentration was key to understanding challenging parts. These accounts show that listening difficulties centered on comprehension, processing speed, and input the learner could not control, consistent with the receptive demands of the skill.

3.2.3. Duration and Number of Questions in the Listening Exam

Students' views on the length and item count of the listening exam ranged across the codes of insufficient time, sufficient time, few questions, sufficient questions, time pressure, and feeling comfortable. Under insufficient time, Ela needed more time to focus on each question and Alp noted that time pressure

made it difficult to think clearly. In contrast, under sufficient time, Derin reported that the allotted time was just enough to listen and respond calmly and Efe believed the duration allowed him to complete all items without stress. Regarding item count, Emir commented under few questions that the exam ended too quickly to reflect his effort, while under sufficient questions Kerem appreciated that the exam did not feel overloaded and Asya explained that the number of questions matched the difficulty level. A related concern was time pressure, as Zeynep described the ticking clock as distracting and Duru frequently checked the time instead of concentrating. Finally, under feeling comfortable, Defne reported feeling calm and found the pacing relaxing and İrem experienced no stress related to timing or question load. These responses indicate that perceptions of exam design were shaped not only by objective features but also by subjective experiences of pressure and preparedness.

3.2.4. Challenges and Facilitators in the Speaking Exam

Students' experiences of the speaking exam clustered into challenges, coded as emotional difficulty, sentence construction difficulty, and word retrieval difficulty, and facilitators, coded as confidence and quick response, topic familiarity, and fluency. Within emotional difficulty, several students reported nervousness that interfered with responding, as Ela stated, *My mind goes blank because of anxiety, even when I study a lot*, and Arda explained, *Even when I know what I want to say, I can't start speaking*. Under sentence construction difficulty, learners struggled to formulate accurate sentences under time pressure, as Derin remarked, *It's hard to form correct sentences quickly*, and Can noted, *Thinking in English and turning my thoughts into speech feels too slow during the exam*. Word retrieval difficulty was also recurrent, as Mert shared, *I can't remember specific words, so I pause too much*, and Alp explained, *I use very basic words because I can't reach more advanced ones at that moment*. Among facilitators, confidence and quick response featured, as Zeynep said, *I'm usually good at thinking on my feet*, and Asya explained, *My speaking reflexes improved thanks to practice*. Topic familiarity helped, as Kerem said, *Knowing the topic helps me focus more on language than ideas*, and Naz added, *It's easier to speak when the subject is something I like*. Fluency reflected personal strengths, as Duru explained, *Once I start speaking, the words flow naturally*, and Baran noted, *I feel in control of my speech from beginning to end*. These accounts show that speaking difficulties centered on performance production and the pressure of being evaluated in real time, in contrast to the comprehension-oriented concerns of listening.

3.2.5. Duration and Number of Questions in the Speaking Exam

Students' perceptions of the time and item count in the speaking exam spanned the codes of insufficient time, sufficient time, few questions, sufficient questions, time pressure, and feeling comfortable. Under insufficient time, Ela explained that the time ran out before she could finish her thoughts and Alp noted that the pressure to finish quickly made it difficult to focus on accuracy. In contrast, under sufficient time, Derin spoke at a comfortable pace without feeling rushed and Melis stated that the limit helped her remain concise and focused. Regarding item count, under few questions Zeynep suggested that more questions would allow a broader evaluation of her skills, while under sufficient questions Kerem described the exam as well-balanced and Asya noted that fewer questions reduced her stress. Under time pressure, Can explained that awareness of the ticking clock negatively affected his fluency and Duru stated that this pressure led her to prioritize finishing over quality. Finally, under feeling comfortable, Baran stated that the format suited his speaking style and İrem reported performing calmly without feeling rushed. These findings indicate that perceptions of speaking exam design were shaped by both objective features and students' emotional responses and planning strategies.

3.2.6. Perceptions of Exam Scores

Students' reflections on their scores revealed interpretations shaped by perceived fairness, expectations, and satisfaction, clustering around the codes of fair, unfair, aligned with expectations, low expectation, high expectation, and satisfaction. Several students described their scores as fair, as Ela stated, *My score matches the effort I put into preparing*, and Mert added, *The grading was fair and the feedback helped me understand my mistakes*. In contrast, some perceived their scores as unfair, as Emir shared, *My result doesn't really reflect my ability, I think anxiety affected me*, and Hasan remarked, *They didn't consider my participation during the term*. A group indicated that their scores aligned with expectations, as Kerem stated, *The result matched my performance, so it wasn't a surprise*, and Asya said, *I expected this result because it reflects my fluency and vocabulary*. Some reported low expectations and were therefore unsurprised, as Alp commented, *I didn't study enough and I was stressed, so the low score wasn't a shock*. Others described high expectations relative to their results, as Zeynep stated, *I thought I had done better than this score*, and Baran noted, *I aimed higher and felt discouraged afterward*. Finally, some expressed satisfaction, as Naz described it as *rewarding after all my effort* and İrem stated, *My score boosted my confidence and made me proud of my progress*. These reflections indicate that scores functioned not only as academic indicators, but also as emotionally meaningful outcomes tied to students' self-perception and expectations.

3.3. Perceived Influence of the Exams on Classroom Participation and Affective Responses

This third theme addresses how students perceived the influence of performance exams on their participation and affective responses during classroom activities. Their accounts clustered around the codes of increased participation, participation out of obligation, no change in participation, and reduced anxiety, reflecting both behavioural and emotional shifts.

A considerable number of students described increased participation, indicating that an upcoming exam encouraged more active involvement in listening and speaking tasks. Ela explained that she began taking classroom tasks more seriously because she recognized their role in preparation, Can reported that after the speaking exam was announced he volunteered more frequently during speaking exercises, and Asya emphasized that the existence of a formal evaluation made her *listen with more purpose* during lessons. In contrast, some students' participation stemmed from external pressure, coded as participation out of obligation. Emir stated that he participated *only because the exam was coming up*, Hasan remarked that he felt *forced to practice*, and Baran described his engagement as *compulsory rather than willing*. A smaller group reported no change in participation, as Arda explained that he *always participates the same way, exam or not*, and Alp noted that his involvement depended primarily on his interest in the topic. Several participants associated the exams with reduced anxiety over time, as repeated exposure to exam-like activities normalized the experience. Kerem explained that practicing with the awareness of an upcoming exam *made it feel less scary when the day came*, Duru stated that the structured activities provided a *clear idea of what to expect*, and İrem reported feeling *less stressed during the actual exam* after regular classroom practice. Taken together, these findings indicate that the exams influenced participation in diverse ways, encouraging purposeful engagement for some students while generating a sense of obligation or leaving habits unchanged for others, and that consistent exposure to exam-related activities could reduce test-related anxiety. The influence of performance exams on participation thus appeared to depend on the interaction between motivational, contextual, and affective factors.

4. Conclusion, Discussion, and Recommendations

This study provides an in-depth picture of how middle school students in one Turkish public school perceive listening and speaking performance exams in English lessons. Overall, students' perceptions revolved around a central tension. Many learners viewed these exams as more authentic and communication-oriented than traditional paper-based tests, yet the same assessments were also experienced as emotionally demanding events marked by uncertainty, anxiety, and concerns about grading and fairness. This duality suggests that performance-based assessment functions not merely as a measurement tool but as a high-impact classroom event that can either foster or constrain learner engagement, depending on how it is designed and implemented. Because these perceptions were formed in a setting where listening and speaking performance exams are a routine, graded component of evaluation rather than an occasional practice, students were reflecting on assessments they encounter repeatedly across the academic year, which gives particular weight to the affective patterns reported here.

Students' reflections indicate that the separate assessment of listening and speaking skills is often perceived as meaningful, as it allows learners to focus on specific abilities and recognize particular strengths and weaknesses. This perception is consistent with the view that communicative assessment should reflect distinct language constructs rather than collapsing multiple skills into a single format (Council of Europe, 2001). At the same time, students' negative reactions reveal that skill separation can also be perceived as a burdensome increase in the total number of exams, leading to heightened fatigue, pressure, and anticipatory anxiety. In a context where these exams contribute directly to grades, the addition of a second skill-specific exam is not experienced as a neutral pedagogical choice but as an additional graded event with real stakes. Such responses align with foreign language anxiety research, which emphasizes that evaluative settings can intensify students' self-consciousness and fear of negative evaluation, particularly during oral tasks where performance is both immediate and public (Horwitz et al., 1986; Köksoy & Deniz, 2025; Öztürk & Gürbüz, 2014). These findings suggest that skill-specific assessment is neither inherently supportive nor inherently threatening. Rather, its impact depends on whether learners perceive the process as predictable, supportive, and aligned with classroom instruction.

Across the interviews, students frequently described listening and speaking exams as more real and more closely connected to actual language use, signalling a strong sense of perceived authenticity. Authenticity is a central principle of communicative language assessment, which aims to approximate real-world language use and elicit meaningful performance (Council of Europe, 2001). However, this perceived authenticity coexisted with concerns about accuracy and fairness. Students emphasized that anxiety, background noise, unclear instructions, or time pressure could negatively affect their performance and distort outcomes. This concern aligns with assessment literature highlighting that performance scores are sensitive to contextual conditions and rating practices, which may threaten valid score interpretation when procedures lack standardization and transparency (Luoma, 2004). According to Luoma (2004), clearly defined rating criteria and consistent assessment procedures are essential for ensuring fair and dependable scoring in speaking assessment. In the present setting, where the exams are administered as part of regular classroom evaluation rather than under centrally standardized conditions, the transparency and consistency of local scoring practices become especially consequential for how fairly students perceive their results. The perception that assessment can have substantial consequences for learners also resonates with washback research showing that high-stakes examinations can intensify anxiety, reduce motivation, and negatively shape language-learning experiences (Özmen, 2011).

The influence of listening and speaking exams extended beyond assessment moments and shaped classroom participation patterns. Many students reported increased participation, suggesting positive

washback effects. When assessment tasks resemble classroom activities, learners are more likely to engage in relevant practice, a widely acknowledged advantage of communicative assessment (Black & Wiliam, 2009; Muñoz & Álvarez, 2010). Nevertheless, some students indicated that their participation was driven primarily by obligation rather than genuine motivation, pointing to a more compliance-oriented form of engagement. This dual pattern reflects a well-documented feature of washback, as the same assessment can affect learners differently depending on individual and contextual factors, including their motivation and relationship to the stakes of the assessment (Muñoz & Álvarez, 2010; Özmen, 2011). In this respect, performance exams operate as powerful classroom levers that shape not only participation but also classroom climate and learners' willingness to take communicative risks.

Students' accounts also highlight the critical role of preparation in shaping exam experiences. Learners who engaged in planned, repeated practice reported greater confidence and more positive perceptions of listening and speaking exams, whereas last-minute or unstructured preparation was associated with uncertainty and stress. These findings suggest that performance exams implicitly demand self-regulation skills that are not evenly distributed among students. In the absence of explicit guidance, preparation itself can become a hidden source of inequality. Interviews further reveal that some students rely on memorization strategies for speaking tasks as a way to manage anxiety. While memorization may provide short-term emotional relief, it risks narrowing the construct of speaking performance by privileging rehearsed output over spontaneous production. This underscores the importance of classroom routines that normalize rehearsal, practice, and feedback as ongoing learning processes rather than emergency responses to upcoming exams (Black & Wiliam, 2009).

The challenges students reported in listening and speaking reveal distinct but interconnected profiles. Listening difficulties were commonly linked to speech rate, unfamiliar vocabulary, accent variation, and cognitive load, whereas facilitators included repetition, clear instructions, and sustained attention. Consistent with some of these findings, Aybirdi et al. (2024) found that learners experienced difficulties particularly with unfamiliar vocabulary and high speech rate, while extensive exposure to authentic listening materials improved listening proficiency and reduced listening anxiety. Speaking difficulties, by contrast, were largely associated with emotional interference, sentence construction under pressure, and problems with lexical retrieval. These difficulties mirror findings in speaking anxiety research showing that anxiety may interfere with lexical retrieval and sentence production during speaking (Horwitz et al., 1986; Öztürk & Gürbüz, 2014). At the same time, students' reports of fluency, topic familiarity, and confidence as facilitating factors suggest that repeated, low-stakes speaking opportunities may help strengthen confidence and fluency over time. Overall, the contrast between the two skills suggests that listening was experienced primarily as a challenge of comprehending and keeping pace with externally controlled input, whereas speaking was experienced as a challenge of producing language under immediate communicative pressure. This distinction indicates that the two skills may require different forms of instructional and assessment support rather than identical treatment.

Students' perceptions of exam design features further illustrate how technical decisions intersect with emotional experience. Even when task duration or the number of questions was perceived as sufficient by some learners, some learners experienced intense pressure due to anxiety and uncertainty. This indicates that design features such as timing and pacing function not only as logistical parameters but also as affective triggers. Research on oral assessment design emphasizes that reliability and fairness depend on clearly defined criteria, consistent procedures, and appropriately designed tasks and instructions (Luoma, 2004). From this perspective, exam design choices play a critical role in determining whether scores can be

meaningfully interpreted as indicators of language competence rather than as reflections of a student's situational anxiety.

Students also reported a wide range of efforts to improve their listening and speaking skills, including repeated listening, audiovisual exposure, self-study routines, mirror practice, vocabulary development, and seeking external support. This diversity reflects a degree of learner agency. However, interviews also indicate that not all strategies were equally structured or effective, and some learners expressed uncertainty about how to improve. These findings point to the instructional value of integrating strategy instruction and modelling practice routines within regular classroom teaching. From a formative assessment standpoint, clear learning goals and feedback can help learners interpret performance information as guidance for improvement rather than as judgement (Black & Wiliam, 2009).

Students' interpretations of scores reveal that grades were experienced not only as indicators of performance but also as reflections of effort, self-worth, and fairness. Learners who perceived a mismatch between their classroom participation and exam scores reported disappointment and mistrust toward grading practices. This finding reinforces a core principle of classroom assessment, namely that students' trust depends on transparent criteria, perceived consistency, and meaningful feedback. Research on teacher assessment competence and language assessment literacy indicates that many teachers feel insufficiently prepared to design and implement reliable assessments for productive skills (Ballıdağ & İnan-Karagül, 2021), a concern that closely parallels students' experiences in the present study and that is especially relevant when, as in this school, the same teachers design, administer, and score the performance exams that shape students' grades.

Taken together, these findings indicate that students in this setting perceive listening and speaking performance exams as valuable in principle but emotionally demanding in practice, and that the way these exams are locally prepared, administered, and scored plays a decisive role in whether they are experienced as fair and supportive. Because the study draws on the perceptions of 20 students in a single public middle school, these conclusions are best understood as context-sensitive insight into one assessment setting rather than as evidence generalizable to public schools more broadly.

Based on the findings, several pedagogical recommendations can be proposed. First, listening and speaking performance assessments should be implemented within more supportive and predictable classroom environments. Since students frequently associated these exams with anxiety, uncertainty, and perceived unfairness, teachers are encouraged to provide clear task formats, transparent scoring criteria, and consistent procedures in advance. Regular low-stakes practice tasks that mirror exam conditions may help normalize performance situations and reduce emotional pressure, allowing students to demonstrate their actual language competence more accurately. Second, structured preparation and strategy instruction should be integrated into everyday teaching. The results showed that learners who engaged in repeated practice, planning, and guided rehearsal reported greater confidence, whereas unstructured preparation increased stress and dissatisfaction. Teachers may therefore model effective listening and speaking strategies, incorporate rehearsal opportunities, and provide formative feedback to support self-regulation. Embedding assessment within ongoing learning processes rather than treating it as an isolated event may contribute to both improved performance and stronger learner agency (Black & Wiliam, 2009).

This study has several limitations that should be acknowledged. First, the findings are based solely on qualitative interview data and reflect students' self-reported perceptions rather than direct measures of performance. Second, the study was conducted with a relatively small sample drawn from a single public

middle school, so the findings are offered as context-sensitive insight rather than as generalizable evidence. Third, the cross-sectional design does not allow for examining changes in students' perceptions or anxiety over time. In light of these limitations, several directions for future research emerge. Future studies should combine perception data with performance scores, classroom observations, and longitudinal evidence to better understand how assessment practices relate to achievement and anxiety over time. Experimental or intervention-based designs that examine the impact of strategy instruction, rubric transparency, or teacher assessment training may provide more robust evidence for effective practices. Comparative research across different contexts and proficiency levels may further clarify how contextual variables shape students' experiences of fairness and stress.

Ethical Statements

This journal adopts the COPE (Committee on Publication Ethics) principles.

Conflict of interest

The authors declare no conflict of interest.

Funding

This study received no financial support from any institution or organization.

Reference verification statement

The authors declare that all works listed in the references are cited within the text and that their accuracy has been verified.

Plagiarism statement

This study has been screened for plagiarism using plagiarism-detection software and has been prepared in accordance with scientific ethical guidelines.

Peer review process

This article has undergone a double-blind peer review process.

Article access and licensing

Authors retain copyright of their work and grant the journal the right of first publication. All articles are licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

References

- Acar, A. (2025). A critical analysis of the 2025 high school English language curriculum of Türkiye. *European Studies in Business and Society Bulletin*, 11(2), 73–93. <http://www.englishscholarsbeyondborders.org/wp-content/uploads/2025/09/Ahmet-Acar.pdf>
- Aybirdi, N., Han, T., & Töngel, P. (2024). The impact of extensive listening on listening anxiety, listening skills and attitude towards English courses. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 6(2), 194–211. <https://doi.org/10.38151/akef.2024.138>
- Ballıdağ, S., & İnan-Karagül, B. (2021). Exploring the language assessment literacy of Turkish in-service EFL teachers. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24(45), 73–92. <https://doi.org/10.31795/baunsobed.909953>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Carcary, M. (2009). The research audit trail: Enhancing trustworthiness in qualitative inquiry. *The Electronic Journal of Business Research Methods*, 7(1), 11–24.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16802fc1bf>
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Findıklı, S., & Büyükkarcı, K. (2024). A comparison of Turkish and European English language teachers' language assessment knowledge levels and perceptions. *International Journal of Assessment Tools in Education*, 11(2), 345–367. <https://doi.org/10.21449/ijate.1360899>
- Gilakjani, A. P., & Sabouri, N. B. (2016). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching*, 9(6), 123–133. <https://doi.org/10.5539/elt.v9n6p123>
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111–128. <https://dergipark.org.tr/en/pub/eltrj/article/308006>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Jamshed, S. (2014). Qualitative research method-interviewing and observation. *Journal of Basic and Clinical Pharmacy*, 5(4), 87–88. <https://doi.org/10.4103/0976-0105.141942>
- Köksoy, B., & Deniz, L. (2025). Foreign language anxiety in middle school students: A case study. *Anadolu University Journal of Education Faculty*, 9(2), 116–148. <https://doi.org/10.34056/aujef.1490320>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733017>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
- Ministry of National Education. (2023). *Millî Eğitim Bakanlığı yazılı ve uygulamalı sınavlar yönergesi*. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2023_10/12115933_MEB_yazili_ve_uygulamali_sinavlar_yonergesi.pdf
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33–49. <https://doi.org/10.1177/0265532209347148>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1–13. <https://doi.org/10.1177/1609406917733847>
- Ölmezer-Öztürk, E., & Aydın, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, 34(3), 602–620. <https://doi.org/10.16986/HUJE.2018043465>
- Özcanlı, N., & Kozikoğlu, İ. (2024). Ortaokul öğrencilerinin İngilizce konuşma öz-yeterlik algıları üzerine bir inceleme. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 26(3), 1030-1046. <https://doi.org/10.32709/akusosbil.1195547>

- Özmen, K. S. (2011). Washback effects of the inter-university foreign language examination on foreign language competences of candidate academics. *Novitas-ROYAL (Research on Youth and Language)*, 5(2), 215–228. <https://izlik.org/JA88JM58DT>
- Öztürk, G., & Gürbüz, N. (2014). Speaking anxiety among Turkish EFL learners: The case at a state university. *Journal of Language and Linguistic Studies*, 10(1), 1–17. <https://izlik.org/JA78JM98YM>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.
- Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.
- Satgun, G., & Aşık, A. (2023). Exploring language assessment knowledge of in-service English language teachers in K-12 context. *Journal of Teacher Education and Lifelong Learning*, 5(2), 610–627. <https://doi.org/10.51535/tell.1321423>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63–75. <https://doi.org/10.3233/EFI-2004-22201>
- Tracy, S. J. (2020). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (2nd ed.). Wiley-Blackwell.
- Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, 6(5), 100–110.