Academic major as a variable in EFL instructors' speaking assessment preferences in preparatory programs

Berk İlhan[1]

[1]Alanya Alaaddin Keykubat University, Antalya, Turkiye / Contact: berk.ilhan@alanya.edu.tr

Abstract

Five departments in Turkiye train English language teachers and EFL instructors at universities. The variation in educational background might affect their assessment practices. This descriptive study investigates whether such differences exist by examining the speaking assessment preferences of 82 EFL instructors working in university preparatory programs. Data were collected via an electronic questionnaire featuring statements on different CEFR-based assessment types. The Kruskal-Wallis test revealed two statistically significant differences in the participants' assessment preferences based on their academic majors. Overall, instructors favor speaking assessments that use language in authentic contexts, apply continuous and formative evaluation, remain objective, and utilize external assessment over self-assessment. They believe such methods foster more reliable and comprehensive measures of language ability. Statistically significant differences were found in criterion-referencing and guided judgment, suggesting a heightened focus on these types in pre-service and in-service teacher education programs to equip instructors with diversified assessment strategies.

Introduction

Teaching speaking is an essential part of teaching a second or foreign language. McDonough (1993) claims that speaking is the skill through which an individual's language proficiency is judged at first sight. On many occasions, language users are only evaluated by their speaking skills. However, in many educational contexts, due to different assessment types, learners either do not feel encouraged to develop their speaking skills or fail to get proper feedback on the weaknesses or strengths of their speaking skills. There is not only one effective way of assessment. According to Heaton (2003), speaking is a crucial ability, although evaluating performance objectively is not always easy. Similarly, according to Brinke et al. (2007), assessments are the primary

element of education; as a result, there are essential aspects to consider while organizing and carrying out speaking instruction and evaluation.

One of these factors is choosing an appropriate assessment type. The Common European Framework of Reference for Languages (CEFR), which guides English teachers in Europe and provides synergy in teaching practices, guides the assessment types (Council of Europe, 2001). From various assessment alternatives, the choice is left to the teachers, and a list of things to consider in speaking assessment is provided in the framework. It is stated that users of the framework should reflect on which type of assessment is appropriate by considering the learners' needs in the context and the appropriateness and feasibility of the assessment type in the educational culture. Further, it is claimed that users of the framework should also consider the extent to which teachers know these assessment techniques and the extent to which they have been trained in using them (Council of Europe, 2001).

In Turkiye, English language teachers in primary and secondary level education and English as Foreign Language (EFL) instructors at the tertiary level graduate from five primary academic majors: English Language Teaching (ELT), English Language Literature (ELL), Linguistics (LNG), American Culture and Literature (ACL), and Translation and Interpreting Studies (TIS). Except for the teachers from the ELT department, the graduates from the other departments must participate in a pedagogical training program to start teaching in primary and secondary-level schools. The program has general pedagogical knowledge classes. To work as an EFL instructor at a university, a Master's Degree (MA) is obligatory, whereas pedagogical training is not. As teachers' educational background might be a determinant factor in their teaching practices, it may also affect their speaking assessment preferences.

To examine which general assessment types EFL instructors prefer in Turkiye for speaking assessment and to reveal whether their speaking assessment preferences change according to their academic majors, this descriptive study has two research questions:

1. What types of assessment do EFL instructors prefer to assess speaking?

2. Is there a difference between their academic majors in Bachelor's degree (BA) and speaking assessment preferences?

Literature Review

The CEFR (Council of Europe, 2001) provides a list of assessment types that can be used to evaluate each language competence individually and as a whole. Since its initial publication, the CEFR Companion Volume has expanded these descriptors—particularly for mediation and plurilingual competence—underscoring the need for updated assessment decisions (Council of Europe, 2020). Teachers must choose the best assessment type considering contextual factors and educational purposes. Recent empirical work links such decision-making directly to teachers' language-assessment-literacy levels (Kremmel & Harding, 2019).

The general assessment types are not only for speaking skills assessment. They can be used to assess four language skills separately or as integrated skills. A systematic review of integrated-skill tasks demonstrates that balanced assessment across modalities significantly enhances communicative accuracy (Zhang et al., 2024). There might be many factors to consider before conducting a speaking assessment and choosing the most appropriate assessment type might be one of them (Sasayama & Norris, 2023). Technology-enhanced formats—for example, online synchronous speaking tasks—are increasingly leveraged to diversify assessment options (Jones et al., 2023). Therefore, this study examines teachers' ideas for assessment types for speaking assessment. The following assessment types in CEFR (Council of Europe, 2001) were included in the study:

Achievement Assessment / Proficiency Assessment: It evaluates what is taught because achievement assessment is used to gauge goal achievement. When assessing achievement, the viewpoint is internal. For example, it may be regarding the material covered in a textbook or the program's syllabus. However, in the assessment of proficiency, the viewpoint is external. It evaluates the abilities of students to apply knowledge or skills they have acquired in the program in real-world circumstances. Recent classroom research in Turkiye indicates that pairing proficiency-oriented speaking tasks with reflective journals can heighten learner engagement (Mutlu, 2025).

Norm-referencing Assessment / Criterion-referencing Assessment: In a norm-referencing examination, learners are ranked according to their scores. It may have a detrimental washback impact because there is competition. In a criterion-referencing assessment, there is no comparison between students and their peers. Instead, the extent to which the students have succeeded in achieving their objectives in the target language is considered. There are no grades in the criterion-referencing evaluation, but learners may receive praise for their work in the form of words and phrases like "good," "well done," and "excellent." Studies show that criterion-referenced tasks supported by transparent rubrics foster positive learner attitudes and reduce competitive anxiety (Fulcher, 2020).

Continuous Assessment / Fixed-point Assessment: The fixed-point assessment can be completed at the end of an academic term or year to determine whether the educational goals have been met. Continuous assessment can be done through projects, presentations, and performance activities during a term or year. Mobile portfolio platforms have made such continuous monitoring of oral development more feasible in regular classrooms (Zhang et al., 2024).

Formative Assessment / Summative Assessment: The practice of gathering data on learning efficiency during a semester is known as formative assessment. It provides teachers with feedback on the effectiveness of their lessons, allowing them to make the required corrections and modifications in light of the assessment's findings. Receiving feedback on the effectiveness of teaching and learning after a semester or academic year is a summative assessment. Meta-analytic evidence confirms that formative oral-feedback cycles can improve both fluency and accuracy (Zhang et al., 2024).

Direct Assessment / Indirect Assessment: While indirect assessment evaluates knowledge or ability through some intermediary activities, direct assessment evaluates skills or knowledge directly through observation. Written dialogues in English or questions with answer options are examples of indirect speaking evaluation. Automated speech-recognition tools are increasingly integrated into indirect tasks to streamline large-scale assessment (Jones et al., 2023).

Performance Assessment / Knowledge Assessment: In performance evaluation, the assessment is carried out by watching a real-world performance. For instance,

35

assessing performance just based on speaking fluency is possible. Regarding knowledge assessment, however, learners' linguistic proficiency and use of control matter more. For instance, learners' usage of various linguistic structures and their capacity to answer questions can be assessed for speaking skills. Task-based tests combining performance and linguistic-knowledge components have demonstrated stronger construct validity in recent validation studies (McNamara et al., 2019).

Subjective Assessment / Objective Assessment: In a subjective assessment, one assessor makes a personal judgment about the merits of an observable performance. There may be various assessors in an objective assessment, and the learners' reactions to the performance may be constrained and regulated. Use of analytic rating scales has been shown to mitigate subjectivity and enhance inter-rater reliability (Isaacs, 2018).

Assessment through Impression / Assessment by Guided Judgment: In contrast to the assessment by guided judgment, where there are standards and a defined assessment process, the evaluation through impression does not have any explicit criteria for examination. The CEFR Companion Volume now advocates guided-judgment approaches to increase transparency in speaking assessment (Council of Europe, 2020).

Holistic Assessment / Analytic Assessment: In a holistic assessment, the performance is evaluated holistically without focusing on various linguistic characteristics, whereas in an analytical assessment, performance sub-skills are evaluated, and the focus may be on multiple linguistic factors. Machine-learning–assisted scoring systems are increasingly paired with analytic rubrics to bolster score reliability (Jones et al., 2023).

Assessment by Others / Self-assessment: The performance is evaluated by an assessor or others; however, in the case of self-assessment, the students evaluate their own or their peers' performance. Knowing one's strengths and flaws is self-assessment. Evidence from CEFR-based self-assessment implementations in Turkish secondary EFL classrooms reveals significant gains in learner reflection and oral proficiency (Yüce & Mirici, 2022).

EFL instructors' educational backgrounds, particularly their academic majors in a BA degree in Turkiye, might be a contextual factor determining their assessment

choices. Parallel findings in East Asian contexts similarly show that disciplinary training influences instructors' preferred assessment modes (Harding & Kremmel, 2019). Therefore, this study was conducted to investigate whether there are different speaking assessment preferences of EFL instructors in Turkiye and whether there are differences among them regarding their academic majors.

A comprehensive literature review revealed some studies on speaking assessment in EFL contexts. However, no studies examine EFL instructors' preferences regarding speaking assessment types. Furthermore, no other studies examine the differences among teachers regarding their educational backgrounds. Recent large-scale surveys continue to highlight this gap, calling for investigation into how CEFR-aligned categories inform university instructors' choices (Mutlu, 2025). The studies on speaking assessment are mainly on teachers' common speaking assessment practices at the university level (Hosseini & Azarnoosh, 2014), speaking assessment practices in primary and secondary education (Matin, 2013), and their relationship with teachers' experience, gender, and education contexts (Oz, 2014), the change in speaking assessment practices in different educational contexts (Cheng et al., 2004), the differences in the theory and practice regarding speaking assessment practices (Kellermeier, 2010), the feelings of the learners and teachers during and after speaking assessment (Hol, 2010), time spared for speaking assessment and practice (Gulluoglu, 2004), and teachers' perceptions for in-class speaking assessment (Thuy & Nga, 2018). None of these studies, as well as the other studies in speaking assessment, have focused on instructors' perceptions of speaking assessment types in the CEFR and the effect of educational background on their preferences. As this study is one of the first examples focusing on these aspects, it might contribute to the field.

Method

The research was conducted under a positivist philosophical stance. Park et al. (2020) state that "studies aligned with positivism focus on identifying explanatory associations or causal relationships through quantitative approaches" (p. 690). There is no intervention; therefore, the research design is descriptive. As Seliger and Shohamy (1989) state, "descriptive research involves a collection of techniques used to specify,

delineate or describe naturally occurring phenomena without experimental manipulation" (p.124). The research aims to describe naturally occurring phenomena, the speaking assessment preferences of EFL instructors and analyze the relationship of the preferences with BA degree majors. It is part of a master's thesis by the researcher (Ilhan, 2017).

Participants

In Turkiye, EFL instructors could be graduates of five academic majors with BA degrees. The MA degree is obligatory to be an instructor at the university; however, pedagogical training is not compulsory. An MA degree can be in the same department as a BA degree. Therefore, each university has EFL instructors with different educational backgrounds, which was the central curiosity behind this study.

Participants in the study were 82 EFL instructors employed by several Turkish universities. They were chosen randomly using a convenience sampling method. In convenience sampling, participants are selected based on accessibility, proximity to the study site, availability at a specific time, and willingness to participate (Dornyei, 2007). The data was gathered using an electronic questionnaire sent to the instructors through their institutional email addresses.

Table 1

Academic Majors of the Participants

	f
1 ELT	42
2 ELL	31
3 TIS	5
4 ACL	2
5 LNG	2
Total	82

The participants were graduates of five BA degree majors. There were 42 graduates from the English Language Teaching (ELT), 31 from the English Language and Literature (ELL), five from the Translation and Interpreting Studies (TIS), two from the Linguistics (LNG) and two from the American Culture and Literature (ACL) departments.

Data Collection

The study used a two-part electronic questionnaire to collect data (see Appendix 1). The first part was for demographic information. In the second part, there were 22 statements for each assessment type. The researcher took the statements directly from the definitions for the assessment types in CEFR (CEFR, 2001). Rather than giving only names, statements were created regarding the definitions, as instructors might not have had the necessary background knowledge for the assessment types. They were changed into a form that the participants could agree or disagree with through 5-point Likert-Scale. After the statements were prepared, they were checked for wording issues by another colleague working at the same institution as the researcher. The questionnaire was sent to 20 instructors who worked at a Turkish University as a pilot study before writing the final questionnaire. Changes were made to the statements to avoid misinterpretation and vagueness, and the final questionnaire (see Appendix 1) was created.

Data Analysis

The analysis was conducted in SPSS 22. Firstly, a descriptive analysis was conducted on central tendency values. As the data deviated from normal distribution, the non-parametric Kruskal-Wallis test, an alternative to the ANOVA, was used to analyze the data. The academic major of the instructors is the study's independent variable, whereas the teachers' preferences for assessment style are the study's dependent variable. Pallant (2010) states that non-parametric analysis techniques can produce more accurate results in small, atypical samples.

Ethical Considerations

The data was collected through an online questionnaire through Google Forms. At the beginning of the questionnaire, the participants were informed about the purpose of the study, how anonymity would be assured, and how the data would be stored. They were told that their participation in the study was voluntary. There was an agreement section for the informed consent form, and participants filled it out if they volunteered to participate. There was no place for the names of participants in the questionnaire; only their academic majors and years of experience in the profession were collected. The anonymity of the participants was assured in that way. The data was kept on the personal computer of the researcher, his thesis supervisor, and the researcher's cloud file. No

ethical harm was expected as the anonymity of the participants was ensured, and the data was kept confidential. Institutional permission was obtained from the university where the study was conducted; however, ethical committee approval was not compulsory when the study was conducted.

Findings

This section will present descriptive analysis and Kruskal-Wallis test findings in the following order. Firstly, speaking assessment preferences will be given regardless of the difference in academic major. Then, the differences among majors revealed through the Kruskal-Wallis test will be provided.

Speaking Assessment Preferences

A descriptive analysis was conducted using SPSS 22 to examine speaking assessment preferences. Measures of central tendency—mean, median—and standard deviation were employed to analyze the data without differentiating between academic majors.

Table 2

Mean Scores of Speaking Assessment Statements

	M	*Mdn*	*SD*
Achievement Assessment	3.24	3	1.00
Proficiency Assessment	4.22	4	0.84
Norm-referencing Assessment	2.10	2	1.08
Criterion-referencing Assessment	3.77	4	1.12
Continuous Assessment	4.00	4	0.86
Fixed-point Assessment	2.45	2	1.11
Formative Assessment	4.09	4	0.83
Summative Assessment	2.37	2	1.09
Indirect Assessment	2.10	2	1.17
Direct Assessment	4.06	4	0.89
Knowledge Assessment	2.84	3	1.16
Performance Assessment	3.76	4	0.86
Subjective Assessment	2.96	3	1.08
Rating on a scale	3.28	3	0.95
Rating on a checklist	3.17	3	1.16
Impression	2.52	2	1.19
Guided judgment	3.98	4	0.92
Objective Assessment	4.05	4	1.05
Holistic Assessment	3.83	4	0.91
Analytic Assessment	3.34	3	1.08
Assessment by others	4.17	4	0.78
Self-Assessment	2.55	3	1.06

According to the data in Table 2, most instructors view speaking assessments as competence tests with a high mean score (4.22). For achievement assessment, the mean score is very close to "neutral" (3.24), indicating that speaking assessment is not solely based on the coursebook and syllabus' contents. The mean score of the norm-referencing is low (2.10), which could mean that teachers disapprove of giving scores, comparing learners with their peers, and putting them in rank order for their speaking performances. They might prefer criterion-referencing assessment more (3.77), which could mean that they like to assess speaking proficiency individually by giving feedback through encouraging words or phrases such as 'that was perfect!', 'you speak fluently,' 'good,' rather than providing scores. A continuous (4.00) and formative assessment (4.09) through collecting different projects, presentations, or tasks during an academic term or year is more favorable than a fixed-point (2.45), summative assessment (2.37) for most participants. There are mediator activities and tasks to assess speaking skills indirectly (2.10), such as written dialogues and question-and-answer type questions. However, the participants in the study prefer direct assessment of speaking skills through direct observation of the performance (4.06). Some participants are neutral about subjective assessment (2.96), and many prefer objective assessment by different assessors and using specific criteria (4.05). Scale-rating and checklist-rating are not assessment types, but they are some of the things to consider in speaking assessment, and they are presented among assessment types in CEFR. Scale and checklist ratings have similar mean scores (3.28 / 3.17); however, more participants support checklist ratings consisting of 'can do' statements or 'yes' or 'no' options for subskills of speaking. Assessment by guided judgment is done through specific criteria and specific procedures for speaking assessment. In contrast, assessment through impression might mean the lack of specific criteria and procedures, and it is more subjective (2.52). Most participants support assessment through guided judgment (3.98). Between the analytic and holistic assessment of speaking, more participants support the holistic assessment (3.83). In this type of assessment, a general score for the learners' overall performance is given, and subskills of speaking are not considered. More participants support assessment by others (4.17) rather than self-assessment (2.55), which means that examiners or instructors could assess speaking instead of learners assessing themselves and their peers.

Differences in Speaking Assessment Preferences According to Majors

Two statistically significant differences were found through the Kruskal-Wallis test.

Table 3

The Kruskal-Wallis Test

	Chi-Square	*df*	*Sig.*
Achievement Assessment	5.085	4	.279
Proficiency Assessment	4.106	4	.392
Norm-referencing Assessment	7.920	4	.095
Criterion-referencing Assessment	**13.481**	**4**	**.009**
Continuous Assessment	1.686	4	.793
Fixed-point Assessment	6.115	4	.191
Formative Assessment	4.008	4	.405
Summative Assessment	9.430	4	.051
Indirect Assessment	6.059	4	.195
Direct Assessment	5.444	4	.245
Knowledge Assessment	4.610	4	.330
Performance Assessment	2.112	4	.715
Subjective Assessment	2.588	4	.629
Rating on a scale	2.425	4	.658
Rating on a checklist	8.569	4	.073
Impression	0.785	4	.940
Guided judgment	**10.109**	**4**	**.039**
Objective Assessment	4.423	4	.352
Holistic Assessment	9.318	4	.054
Analytic Assessment	1.612	4	.807
Assessment by others	4.115	4	.391
Self-Assessment	.076	4	.999

It is apparent in Table 3 that there were two statistically significant differences regarding academic majors. One of the differences was for the criterion-referencing assessment (.009), and the other was for guided judgment (.039).

Table 4

Central Tendency for Criterion-Referencing Assessment

	N	*M Rank*	*Mdn*
1 ELL	30	32.92	4
2 ELT	42	42.73	4
3 LNG	2	70.50	5
4 ACL	2	36.75	3
5 TIS	5	64.90	5
Total	81		4

The Kruskal-Wallis test revealed a statistically significant difference between participants' opinions and their majors for criterion-referencing evaluation. x^2 (4, n= 81) = 13.48, p =.009 (ELL; n= 30, ELT; n= 42, LNG; n= 2, ACL; n= 2, TIS; n= 5). With a median score of 5, two academic majors—LNG and TIS—provided the most support.

The subjects with the second-highest median scores were ELL and ELT (*Mdn* = 4). With the lowest median score for the criterion relating to speaking evaluation, ACL had the lowest mean score. (*Mdn* = 3). The pairwise comparisons through the Dunn test revealed that the biggest statistically significant difference was among instructors with ELL and TIS majors (*p* =.030) for criterion-referencing assessment.

Table 5

Central Tendency for Guided Judgment

	N	*M Rank*	*Mdn*
1 ELL	31	34.61	4
2 ELT	42	42.69	4
3 LNG	2	69.00	5
4 ACL	2	69.00	5
5 TIS	5	52.20	5
Total	82		4

Majors, LNG, and ACL achieve the highest mean scores (69), as Table 5 indicates. Participants from these majors could be prone to using guided judgment to evaluate the speech. With a mean score of 52.20, TIS comes in second behind the two majors. The lowest mean scores are for ELT and ELL majors, at 42.69 for the former and 34.61 for the latter. 'Agree' or 'Totally agree' were the questionnaire's median results for all the major responses. It can be assumed that everyone who participated, regardless of their majors, agrees that evaluation should be done using criteria. (ELL, *n*= 31, ELT, *n*= 42, LNG, *n*= 2, ACL, *n*= 2, TIS, *n*= 5); x^2 (4, *n*= 82) = 10.10, *p* =.039.) The participants who selected "neutral" may base their evaluation of speaking on their impression. The pairwise comparisons through the Dunn test revealed that the biggest statistically significant differences were among instructors with ELL and LNG majors (*p* =.036) and with ELL and ACL majors (*p* =.036) for guided-judgment.

Discussion

According to the results, it may be inferred that EFL instructors believe speaking assessments reflect what students can do or know about using language in the real world. The findings align with the findings of the recent research. Classroom surveys show that over three-quarters of secondary-school EFL teachers now design speaking tasks that deliberately mirror authentic communicative events such as job interviews, service

encounters, and academic presentations (Swaie & Algazo, 2023). Incorporating real-world activities rather than just those found in the course book or syllabus can be preferable (Herrington & Oliver, 2000). Assessment of speaking abilities should be continuous (Cheng et al., 2004). A large-scale systematic review of formative assessment in K-12 EFL contexts concluded that ongoing, low-stakes checks of oral performance consistently boost achievement and motivation compared with single end-term tests (Zhang et al., 2024). Evaluating students all at once at the end of a term may not be as beneficial. It might be preferable to assess them through several cumulative activities, such as projects or tasks, over the academic term or year (Zhou, 2013). As a result, formative rather than summative evaluation should be used as suggested by another research (Ismail et al., 2022; Sirianansopa, 2024). This trend is echoed in recent regional studies that document a steady shift from test-driven practices toward assessment for learning in EFL classrooms (Swaie & Algazo, 2023). According to the participants' responses, speaking assessments provide learners with ongoing feedback to identify the areas of difficulty. High levels of agreement on direct evaluation could indicate that instructors prefer to grade speaking through direct performance observation. Current evidence using gauge-repeatability and reproducibility analysis demonstrates that multi-rater, performance-based speaking assessments markedly increase inter-rater reliability and scoring fairness (Sureeyatanapas et al., 2024). It could be inferred that using intermediary activities will not be as efficient as direct evaluation. Scoring of the performances by various assessors might be more effective and objective. It denotes that instructors favor doing an objective speaking evaluation instead of performing an impressionistic and arbitrary assessment. Assessments made by learners' peers or themselves were considered insufficient, as teachers and examiners are more knowledgeable in assessing speaking skills. Nevertheless, controlled interventions reveal that well-scaffolded peer and self-assessment can foster self-regulated learning and critical thinking without compromising score accuracy (Kumar et al., 2023). Further studies could focus on differences between teacher, self, or peer assessment types.

Two statistically significant differences were found for two assessment types, criterion-referencing assessment and assessment through guided judgment, among five academic majors. Criterion referencing speaking assessment had a high mean score in overall speaking assessment preferences. However, the differences in the criterion-

referenced speaking assessment were found to be statistically significant among majors. The highest consensus came from TIS and LNG majors, with a median value of 5, followed by ELT and ELL. The minor support was from the ACL major. It can be concluded that there might be differences in instructors' thoughts on assessing learners' speaking skills through scores and comparing learners' speaking ability with their classmates' speaking ability. Instructors with TIS and LNG majors might think that scores do not have to be given for learners' speaking skills, and only reinforcement words or phrases, such as 'good,' 'perfect,' and 'well done,' can be enough. They might think that learners' speaking ability must be judged only by considering their proficiency without any comparison with their peers. Instructors with an ACL major neither agree nor disagree with this statement. Although teachers with ELT and ELL majors concur with the idea, they might think reinforcement words and phrases might not be enough, instead of giving scores. Some instructors with ELT and ELL majors might prefer to provide scores for speaking performance and compare the speaking abilities of individual learners with their peers. Comparable discipline-linked divergences in assessment orientation have been documented in other tertiary EFL programs, where language-focused departments favour qualitative feedback over numerical grades (Phung & Michell, 2022). This study has presented similar findings.

Participants' suggestions for directed judgment in speaking assessment suggested a statistically significant difference, such as evaluating speaking using particular criteria. At the same time, some participants with ELT and ELL academic majors did not strongly agree with the assessment type; LNG, ACL, and TIS majors supported using specific criteria when assessing learners. Assessing by impression was the reverse of the guided judgment statement. It may be inferred from the mean scores of the academic majors that some instructors with ELL and ELT educational backgrounds might prefer to assess speaking abilities based on impressions. Research on teacher cognition indicates that rubric-guided analytical judgments generally yield more trustworthy results than quick holistic 'gestalt' impressions, reinforcing the value of explicit criteria in speaking assessment (Phung & Michell, 2022). Therefore, assessing through guided judgment could be significant in speaking assessment.

Conclusion

This study sought to identify the variations in speaking assessment preferences among English teachers employed by Turkish universities' English preparatory programs. Turkiye has five academic majors that educate language teachers; thus, these practices were examined to see variations in the majors' preferences for speaking assessment types. The following responses to the research questions can be provided considering the findings.:

1. What types of assessment do EFL instructors prefer to assess speaking?

EFL instructors at the tertiary level in Turkiye might prefer a proficiency, continuous, formative, direct, and objective assessment for speaking skills. They prefer assessment by others, not self-assessment by learners themselves. They do not choose norm-referencing, fixed-point, summative, or indirect assessment types for speaking skills. Instead of assessing through impression, they might prefer to assess by guided judgment.

2. Is there a difference between their academic majors in Bachelor's degree (BA) and speaking assessment preferences?

Two statistically significant differences were found among academic majors. One was for guided judgment, and the other was for criterion-referencing assessment.

Guided judgment means assessing speaking through specific criteria. EFL instructors with academic majors, LNG, ACL, and TIS prefer to assess speaking skills through measures. The opposite of guided judgment was assessment through impression, which could mean that teachers who do not choose to assess with guided judgment might prefer assessment through impression. The results suggested that EFL instructors with an ELL background might prefer guided judgment less than those with an ELT major. The instructors with the ELL major might sometimes assess through impression more than those with the ELT major. It could be essential to provide training for instructors with ELL to assess speaking through guided judgment.

Criterion-referencing assessment is "assessing speaking to give feedback on where a learner is, irrespective of their peers' ability. This feedback can only be words such as 'well done!' or 'good job!' without any numerical scores. The opposite is norm-

referencing assessment, in which scores are given, and learners are ranked accordingly. The biggest and statistically significant difference was for the criterion-referencing assessment in the study. The results revealed that instructors with LNG and TIS majors might most use criterion-referencing assessments most. They might prioritize speaking assessment functioning as feedback on each learner's ability and as a reinforcement. Instructors with majors in ELL and ACL have the lowest mean score for criterion-referencing assessments, which could mean they might tend to do more norm-referencing assessments by giving scores to learners and putting them in the rank order. As findings revealed, instructors with an ELT major might prefer to use both assessment types for speaking assessment.

It is important to note differences to ensure collaboration and harmony among instructors from various educational backgrounds. The study has revealed two differences in the Turkish context: guided judgment and criterion-referencing assessment. Both are important for an effective speaking assessment; therefore, the differences could be lowered through in-service training or discussions with colleagues. Furthermore, especially in oral exams with specific criteria to assess, instructors with different educational backgrounds could be paired together to avoid harming the assessment process due to different ideas.

Implications, Limitations of the Study, and Suggestions for Further Research

The study is significant because it presents the speaking assessment choices of EFL instructors at the tertiary level in Turkiye. Furthermore, it sheds light on the differences among instructors regarding their academic majors and speaking assessment choices. It is one of the first studies with this focus in the Turkish EFL context.

The study might have implications for pedagogy education. As there are five different majors for EFL instructors in Turkiye, knowing general preferences and differences might be valuable while planning pedagogical education. More focus could be given to norm-referencing and criterion-referencing assessments and guided judgment and assessment through the impression in pedagogical education programs.

The study was conducted with 82 participants, and it is descriptive. More studies with more participants are needed. The data was collected through an electronic

questionnaire, and the response rate was low, as it is the limitation of randomly e-questionnaires for randomly chosen participants. The participants were not normally distributed in the study; therefore, a non-parametric analysis was conducted. Analyzing the data by parametric tests may give more detailed results. More studies with more participants from each academic major and through a parametric analysis might be needed. Reasons for the differences among EFL instructors regarding their majors can be studied further through a qualitative or mixed-method study. The participants were EFL instructors at the tertiary level. As the difference could be observed with language teachers at primary and secondary levels, similar studies could be conducted in these contexts with English teachers.

Ethics Committee Permission Information

This study is a small part of a Master's thesis called "A study on oral assessment practices in English classes applied by language instructors" (Ilhan, 2017). Informed consent was obtained from all participants, and all personal identifiers were removed or anonymized to protect confidentiality. Participation was entirely voluntary. No deception was employed, and no foreseeable risks beyond those encountered in everyday life were imposed. Ethical approval was not sought for the present study because, at the time of the study, it was not obligatory to get ethical committee.

Acknowledgment

References

Brinke, D.J., Bruggen, J.V., Harmans, H., Burgers, J., Giesbers, B., Koper, R., & Latour, I. (2007). Modeling assessment for the re-use of traditional and new types of assessment. *Computers in Human Behavior, 23* (6), 2721-2741. https://doi.org/10.1016/j.chb.2006.08.009

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment (Companion volume)*. Council of Europe Publishing.

Cheng, L., Rogers, T. & Hu, H. (2004). ESL/ EFL teachers' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, *21* (3), 360-389. https://doi.org/10.1191/0265532204lt288oa

Chow, A., & Leung, P. (2011). Assessment for learning in language classrooms. In R. Berry and B. Adamson (Eds.), *Assessment reform in education, education in the Asia-Pacific region: Issues, concerns, and prospects* (pp. 135-154). Springer.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Fulcher, G. (2020). Operationalising assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: Theory and practice* (pp. 1–22). Cambridge Scholars Publishing.

Gulluoglu, O. (2004). *Attıtudes towards testing speaking at Gazi University preparatory school of English and suggested speaking tests* [Unpublished master thesis]. Gazi University.

Heaton, J.B. (2003). *Writing English language tests*. Longman.

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research & Development, 48*(3), 23-48. https://doi.org/10.1007/BF02319856

Hol, D. (2010). *Perceptions and attitudes of the English Language teachers and preparatory learners towards testing speaking communicatively in the School of Foreign Languages at Pamukkale University* [Unpublished master thesis]. Pamukkale University.

Hosseini, A.S. & Azarnoosh, M. (2014). Iranian EFL teacher's speaking Assessment practices: purposes, methods, procedures. *Procedia - Social and Behavioral Sciences*, *98*. 653 – 658. https://doi.org/10.1016/j.sbspro.2014.03.464

Ilhan, B. (2017). *A study on oral assessment practices in English classes applied by language instructors* [Unpublished master thesis]. Ataturk University.

Isaacs, T. (2018). Shifting sands in second-language pronunciation teaching and assessment research and practice. *Language Assessment Quarterly, 15*(3), 273–293. https://doi.org/10.1080/15434303.2018.1472264

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia, 12*(1), 40. https://doi.org/10.1186/s40468-022-00191-4

Jones, D. M., Cheng, L., & Tweedie, G. (2023). Automated scoring of speaking and writing: Starting to hit its stride. *Canadian Journal of Learning and Technology, 48*(3), 1–22. https://doi.org/10.21432/cjlt28241

Kalayci, S. (2016). *SPSS uygulamalı çok değişkenli istatistik teknikleri (Multivariate statistical techniques with spss applications).* Asil Publishing. ISBN: 9759091143

Kellermeier, G. M. (2010). *Foreign language speaking assessment practices in Florida middle and high schools* [Unpublished doctorate dissertation]. University of Central Florida.

Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language-assessment literacy across stakeholder groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly, 16*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self vs. peer assessment activities in EFL-speaking classes: Impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia, 13*(36). https://doi.org/10.1186/s40468-023-00251-3

Matin, Z. N. (2013). Speaking assessment at secondary and higher secondary levels and students' deficiency in speaking skill: A study to find interdependence. *Stamford Journal of English, 7.* 234–251. https://doi.org/10.3329/sje.v7i0.14476

McDonough, J. (1993). *Materials and methods in ELT: a teacher's guide*. Blackwell Publishing.

McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice and language assessment: The role of measurement.* Oxford University Press.

Mutlu, A. (2025). Turkish EFL teachers' beliefs, practices and challenges on assessment. *Manisa Celal Bayar University International Journal of English Language Studies, 4*(1), 1–19.

Oz, H. (2014). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research, 5*(4), 775-785. https://doi.org/10.4304/jltr.5.4.775-785

Pallant, J. (2010). *SPSS survival manual*. McGraw-Hill.

Park, Y.S., Konge, L. & Artino, A. R. (2020). The positivism paradigm of research. *Academic Medicine, 95*(5), 690-694. https://doi.org/10.1097/ACM.0000000000003093

Phung, D. V., & Michell, M. (2022). Inside teacher assessment decision-making: From judgement gestalts to assessment pathways. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.830311

Sasayama, S. & Norris, J.M. (2023). Designing speaking tasks for different assessment goals: The complex relationship between cognitive task complexity, language performance, and task accomplishment. *Journal of Task-Based Language Teaching and Learning, 2* (2), 184-217. https://doi.org/10.1075/task.21020.sas

Seliger, H.W, & Shohamy E. (1989). *Second language research methods.* Oxford University Press.

Siriansopa, K. (2024). Evaluating students' learning achievements using the formative assessment technique: A retrospective study. *BMC Medical Education, 24*, 1373. https://doi.org/10.1186/s12909-024-06347-5

Sureeyatanapas, P., Panitanarak, U., Kraisriwattana, J., & Sarootyanapat, P. & O'Connell, D. (2024). The analysis of marking reliability through the approach of gauge repeatability and reproducibility (GR&R) study: A case of English-speaking test. *Language Testing in Asia, 14*(1). https://doi.org/10.1186/s40468-023-00271-z

Swaie, M., & Algazo, M. (2023). Assessment purposes and methods used by EFL teachers in secondary schools in Jordan. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1192754

Thuy, N.H.H., & Nga, T.T.T. (2018). An investigation into EFL teachers' perceptions of in-class English speaking assessment. *VNU Journal of Foreign Studies, 34*(2). https://doi.org/10.25073/2525-2445/vnufs.4251

Wu, Y. (2013, 3rd July). *Bridging assessment for learning to self-regulation in Chinese tertiary EFL writing classrooms* [Paper presentation]. The 35th Language Testing and Research Colloquium, Seoul.

Yüce, E., & Mirici, İ. H. (2022). Self-assessment in EFL classes of secondary education in Türkiye: Common European Framework of Reference for Languages (CEFR)-based implementations. *Pegem Journal of Education and Instruction, 13*(1), 349–359. https://doi.org/10.47750/pegegog.13.01.38

Zhang, H., Ge, S., & Saad, M. R. B. M. (2024). Formative assessment in K-12 English as a foreign language education: A systematic review. *Heliyon, 10*(10). https://doi.org/10.1016/j.heliyon.2024.e31367

Zhou, J. (2013, 3rd July). *Reconciliation between assessment for learning and assessment of learning in Chinese award-winning teachers' EFL classrooms* [Paper presentation]. The 35th Language Testing and Research Colloquium, Seoul.

Appendix 1

PRACTICES OF ORAL ASSESSMENT AT TERTIARY LEVEL AT TURKISH UNIVERSITIES

This survey aims to uncover common oral assessment methods conducted formally or informally in language classes at universities and whether there is a relationship between practices and academic majors of language instructors.
Thank you for your participation.

1. SECTION: DEMOGRAPHIC INFORMATION

1. Gender: Male **(   )** Female **(   )**

2. What is your academic major? ;
(   ) Department of English Language and Literature
(   ) Department of English Language Teaching
(   ) Department of Linguistics
(   ) Department of American Culture and Literature
(   ) Department of Translation and Interpreting Studies
(   ) Others (Please specify; _____)

3. How long have you been teaching? ; _____ years.

4. What is your age?; _____

5. Is your institution public or private? ; (   ) Public **(   )** Private

6. In which degree do you have classes? ;

(   ) Preparation classes
(   ) Undergraduate classes
(   ) Graduate classes

7. What are your students' levels?

(   ) Beginner **(   )** Elementary **(   )** Pre-intermediate **(   )** Intermediate **(   )** Upper-Intermediate

2. SPEAKING ASSESSMENT PERCEPTION

1. Considering your oral assessment practices, which of these statements do you agree or disagree with? Choose one from **'strongly disagree'** to **'strongly agree'**;

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Speaking assessment is what students can do in activities related to the syllabus course book. It is a kind of feedback for instruction. **(Achievement Assessment)**	( )	( )	( )	( )	( )
Speaking assessment is what students can do/know when applying the language in the real world. **(Proficiency Assessment)**	( )	( )	( )	( )	( )
When providing feedback for speaking, it is better to put students into a rank order and compare them with others in class by giving numerical scores or percentages. **(Norm-Referencing (NR))**	( )	( )	( )	( )	( )
Assessing speaking is essential to give feedback on where a student is, irrespective of their peers' ability. Feedback can be just words such as 'well done!' or 'good job!' without any numerical scores. **(Criterion Referencing (CR))**	( )	( )	( )	( )	( )
Speaking assessment is better done cumulatively by collecting activities such as projects and tasks applied successively during a term. **(Continuous Assessment)**	( )	( )	( )	( )	( )
It is better and more practical to assess speaking with an activity or a task at the end of the term on a particular day. **(Fixed-point Assessment)**	( )	( )	( )	( )	( )
Speaking assessment is ongoing feedback on the efficiency of instruction for teachers and feedback for students to be aware of their weaknesses. **(Formative Assessment)**	( )	( )	( )	( )	( )
Speaking assessment is giving a score for students' speaking competence at the end of a term. **(Summative Assessment)**	( )	( )	( )	( )	( )
Speaking skills can be assessed using mediator test items such as written dialogue completion or multiple-choice tests. **(Indirect Assessment)**	( )	( )	( )	( )	( )
Assessing speaking skills can be done by observing students' performances directly. **(Direct Assessment)**	( )	( )	( )	( )	( )

Statement					
Different test items, even written ones such as filling in the blanks or multiple-choice, can be used to assess speaking and provide evidence for the extent of students' linguistic knowledge and control. **(Knowledge Assessment)**	( )	( )	( )	( )	( )
While assessing speaking, students should be required to provide some samples of the target language to assess them directly. **(Performance Assessment)**	( )	( )	( )	( )	( )
The class teacher can subjectively judge the quality of speaking performance in an assessment. **(Subjective Assessment)**	( )	( )	( )	( )	( )
Assessing speaking skills is judging that students are at a particular level or a band on a scale from 'Very Poor' to 'Very Strong' according to their performances. **(Rating on a Scale)**	( )	( )	( )	( )	( )
Using checklists with 'can do' statements and 'yes' or 'no' tick boxes is ideal while assessing speaking skills. **(Rating on a Checklist)**	( )	( )	( )	( )	( )
Assessing speaking can be done without any specific criteria. Teachers can use their impressions to give a score. **(Impression)**	( )	( )	( )	( )	( )
Assessing speaking can be done by using specific criteria. **(Guided Judgment)**	( )	( )	( )	( )	( )
Scoring is better done by having different assessors objectively score the same performance. **(Objective Assessment)**					
Speaking skills can be assessed holistically by intuitively weighing different aspects and competencies of language. **(Holistic Assessment)**	( )	( )	( )	( )	( )
Speaking skills can be assessed analytically by considering each sub-skill or competence of speaking apart. **(Analytic Assessment)**	( )	( )	( )	( )	( )
Assessing speaking can be done by examiners and teachers. **(Assessment by Others)**	( )	( )	( )	( )	( )
Assessing speaking skills can be done by students themselves or their peers. **(Self-Assessment)**	( )	( )	( )	( )	( )