

Developing a Speaking Assessment Rubric and Strategies for Enhancing Oral Communication in EFL Classrooms

D.M. Akizhanova¹, A.U. Serikbayeva^{2*}

²PhD Student, Al-Farabi Kazakh National University

¹PhD, Head of Translation Theory and Practice Department, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100600171>

Received: 16 May 2026; Accepted: 21 May 2026; Published: 19 June 2026

ABSTRACT

Background: This study aims to develop pilot a contextually appropriate speaking assessment and to identify effective pedagogical strategies for fostering oral communication skills in EFL classrooms.

Methods: A quantitative pilot study was conducted involving 55 students from three 5th grade classes. Participants completed two comparable speaking assignments. Descriptive statistics and Pearson correlation analysis were employed to examine score consistency and the reliability of the rubric across tasks.

Results: The results demonstrated that students' performance across the two tasks was relatively stable, with similar mean scores and moderate variability. A moderate positive correlation ($r = 0.65$) was observed between Task 1 and Task 2, indicating acceptable reliability of the rubric in the pilot phase. However, some individual score fluctuations emphasized potential task-related variability and highlighted the importance of using multiple tasks in speaking assessment.

Conclusion: The findings demonstrate that integrating assessment with instructional speaking strategies significantly enhances learners' fluency, confidence, and communicative effectiveness. The study contributes to the field of language assessment by providing a locally developed rubric that integrates psychometric principles with practical classroom application. The proposed rubric and strategies offer valuable implications for teachers aiming to enhance oral communication skills in EFL contexts.

Keywords: speaking assessment, rubric development, reliability, EFL, Kazakhstan, psychometric evaluation

INTRODUCTION

Assessment is an integral part of any effective educational system. As Gronlund, (1998) mentioned assessment plays a vital role in instruction, and its main goal is to improve learning. Teachers can classify, categorize and grade their students, provide feedback, and adjust their lesson plans by using proper assessment.

According to Andrade (2000) rubrics are effective assessment tools that provide clear criteria for assessment however, it should be used for student development, not just for grading. Rubrics help students comprehend what has to be done, provide more detailed feedback, and boost motivation.

Nowadays, communicative competence plays a pivotal role in teaching English. In Kazakhstan, the UNT, i.e. high-stakes exam, only reading and grammar skills are considered in the English section. In addition, the assessment system in Kazakhstan is often test-oriented, and speaking skills are not sufficiently assessed.

In today's developing society increasingly prioritizes real-life communication, the ability to assess speaking performance accurately and fairly has become a major concern in language teaching processes. Despite

the widespread implementation of speaking rubrics in English language classrooms, many of these assessment tools are employed without sufficient empirical evidence supporting their validity and reliability. In particular, there is limited methodological justification demonstrating whether rubric criteria consistently measure speaking proficiency and whether assessment results remain stable across different raters. The aim of this study is to develop and empirically validate an analytical speaking assessment rubric for English language learners by examining its reliability, construct validity, and rater consistency using advanced psychometric methods. In doing so this study addresses the following research questions:

1. To what extent does the speaking assessment rubric demonstrate internal consistency as measured by Cronbach's alpha?
2. Does the rubric exhibit acceptable inter-rater reliability and acceptable construct validity based on CFA results?
3. What factor structure emerges from exploratory factor analysis of the rubric criteria?

As Messick (1995) noted, psychometric assessment of speech should be based on the theory of construct validity. However, previous studies have not considered the development of a speaking rubric and its full psychometric validation (EFA, CFA, item stability) in a school context, and such comprehensive studies are lacking, especially in the case of Kazakhstan. Speaking ability is increasingly recognized as a core outcome of communicative language teaching. Despite this emphasis, assessment practices often fail to support speaking development, prioritizing summative outcomes rather than formative processes. In many EFL contexts, including Kazakhstan, learners demonstrate reluctance to speak due to fear of errors, lack of confidence, and unclear assessment expectations.

LITERATURE REVIEW

Nowadays, the development of oral communication in English is one of the most crucial areas of forming students' language competence, and speaking skills require not only linguistic knowledge, but also free expression, interaction, and communicative competence.

The issues of speaking assessment and the development of oral communication have been widely studied by foreign and domestic researchers. In particular, a number of studies show that speaking assessment rubrics allow for the systematic assessment of students' oral language.

According to Andrade (1999) Rubrics are important not only for grading, but also as a teaching tool because it shows to the student what is right and wrong. He also claims that rubrics develop their critical thinking and students who use rubrics perform better.

Research has shown that regular feedback during class significantly increases students' learning outcomes and helps them develop self-assessment skills (Black & Wiliam, 1998).

Tuan (2012) revealed that the analytical assessment method effectively developed students' speaking skills. The study found that the analytical group students performed significantly better.

Additionally, Yalçın-Çolakoğlu and Selçuk (2019) show that in speaking exams, assessors rely not only on the rubric, but also on additional factors. They also highlight additional factors such as body language, eye contact and confidence, but they also state that these factors pose a threat to validity.

Miyazoe (2025) found that the P-score model increases motivation by showing students' learning progress. In her study, Miyazoe also noted that the Rater is biased, while the Rubric is still subjective and raised the issue of validity.

Liu et al. (2025) investigates an AI-based speaking assessment tool (EAP Talk). The study examines the accuracy, validity, and effectiveness of this tool in teaching. The research work emphasizes that AI is not a complete solution and should be used as an aid, and that human factors are important in speaking assessment.

Mawalim et al., (2025) claims assessing speaking requires looking at the entire process of speaking, not just grammar, so the author proposes a multimodal AI + explainable model. This study presents an AI model that uses multimodal data to accurately and interpretably assess speaking skills, demonstrating the importance of aspects of speech such as prosody and interaction.

Rubrics also improve student learning, make assessments more accurate and fair, and make assessments more reliable for teachers. Further research has demonstrated that Student performance is often higher when rubrics are used also rubrics useful and understandable (Andrade & Reddy, 2010)

Regarding assessment, in Kazakhstan, Serikbayeva et al. (2026) suggest that the assessment of intercultural communicative competence (ICC) should be considered alongside language competence. In this regard, the use of psychometrically validated assessment tools, especially rubrics, can increase the accuracy, fairness, and effectiveness of speaking assessment in EFL classrooms.

Moreover, Tleuov, (2025) Teaching speaking skills in EFL classrooms in Kazakhstan shows that it is limited by conflicts between teachers' beliefs and the examination-oriented system. He also emphasizes Teaching Speaking: depends not only on the method but also on the context (exam, policy, time).

Moreover, limitations of speaking and assessment have been studied by Kazakhstani scholars as Tleuov (2025), research on creation and validation of speaking evaluation rubrics in secondary schools EFL contexts remain underexplored.

According to Begimbetova et.al. (2023) the criterion-based assessment system is widely used in Kazakhstan and is the only objective form of assessment. However, such approaches mainly focus on general assessment practices and lack psychometrically validated speaking-specific rubrics.

Smakova and Paulsrud (2020) examined the perspectives of EFL teachers in Kazakhstan on intercultural communicative competence (ICC). However, its implementation in classroom practice remains inconsistent due to methodological and assessment-related challenges.

Moreover, recent research highlights the importance of formative speaking activities, such as task-based interaction, situational dialogues, and reflective self-assessment, in increasing learners' willingness to communicate. However, limited studies address the integration of speaking assessment rubrics with explicit speaking development strategies, which constitutes the research gap addressed in this study.

METHODOLOGY

Research design

A mixed-methods research design was employed, combining qualitative classroom observations with quantitative rubric-based speaking assessments. The study was conducted in a secondary school EFL context.

Participants

Participants included EFL learners from comprehensive school №171. A total of fifty-five students (N= 55) participated in the study. Participants were randomly selected from fifth graders. The study involved 55% girls and 45% boys aged 11-12.

Instruments

For the study, a two-part Speaking test was used, adapted to the specific context. The first part answers basic questions, for example Task 1. Speak about your weekend. (Where did you go? Who were you with?, What did you do?) Students talk about how their weekend was. The student discusses a particular picture in the second section. This oral exam is unique in that it is tailored to the Kazakh environment. In the modified test, the student is requested to discuss the capital city of Astana if he has read about the Eiffel Tower in the textbook.

Additionally, he discusses Kokpar in the oral exam, which is specifically tailored to the Kazakh environment, even if the textbook discusses a variety of sports. The oral exam consists of two parts, each lasting five minutes.

An analytical speaking rubric was developed consisting of five criteria: content relevance, lexical range, grammatical accuracy, pronunciation intelligibility, and fluency. Each criterion was assessed on a four-level performance scale. To support speaking performance, the following instructional strategies were systematically implemented: low-stakes speaking tasks; situational role-play activities; peer-supported dialogue construction; self-assessment using rubric descriptors.

Procedure

Students took a contextualized oral test based on a specially developed rubric. The assessment rubric is based on four specifically designed criteria: fluency, pronunciation, grammar, vocabulary. In a session that met three times a week for a week, fifty-five pupils took the test.

Criteria	Excellent 9-10 points	Good 7-8 points	Satisfactory 5-6 points	Needs Improvement 1-5 points
Fluency	Speaks smoothly with no hesitation; ideas are well connected	Minor hesitation; generally fluent	Frequent pauses; limited flow	Very hesitant; speech is fragmented
Pronunciation	Clear and accurate pronunciation; easy to understand	Minor pronunciation errors	Noticeable errors but meaning clear	Frequent errors; difficult to understand
Grammar	Uses correct grammatical structures consistently	Few grammatical errors	Some errors affecting clarity	Frequent errors; meaning unclear
Vocabulary	Uses a wide range of appropriate vocabulary	Adequate vocabulary with minor repetition	Limited vocabulary; frequent repetition	Very basic vocabulary; inappropriate use

Table 1. Speaking Assessment Rubric

Data analysis

The data collected during the study were analyzed using the Jamovi Software program version 2.6.45. First, descriptive statistics (mean, standard deviation) were calculated, and the students' speaking results were described at a general level. In order to determine the reliability of the assessment tools, the internal consistency coefficient (Cronbach's alpha) was used. Furthermore a t-test was used to determine the differences in the results of students. This method allowed us to check the statistical significance of the grades obtained on the basis of the rubric. Where necessary, factor analysis (CFA) was used to check the structure of the criteria, which was aimed at determining the structural validity of the assessment tools.

RESULTS

Variable	Mean (M)	SD	Min	Max
Fluency	2.17	0.75	1	3
Pronunciation	2.20	0.65	1	3
Grammar	1.62	0.49	1	2
Vocabulary	1.67	0.47	1	2
Total Score	7.67	1.36	5	10

Table 1. Descriptive Statistics for Speaking Tasks.

This table shows the descriptive statistics of the students' speaking skills. The overall result shows that the students' level is average. According to the results, fluency M=2.17 is close to the average level, and

pronunciation M=2.20. Pronunciation is well developed skill M=2.20 whereas the last two skills grammar M=1.62 and vocabulary M= 1.67are the weaker ones.

Variable	Task 1 M (SD)	Task 2 M (SD)
Fluency	2.02 (0.80)	2.33 (0.67)
Pronunciation	2.18 (0.58)	2.22 (0.71)
Grammar	1.65 (0.48)	1.58 (0.50)
Vocabulary	1.69 (0.47)	1.76 (0.43)
Total Score	7.55 (1.44)	7.80 (1.27)

Table 2. Comparison of Task 1 and Task 2.

It is observed that students spoke fluently in Task 2, showing a significantly better result of 2.02 in Fluency for the first task and 2.33 in the second task. It was also found that Pronunciation remained at a stable level with indicators of 2.18 and 2.22. The Grammar score is between 1.65 and 1.58, which means that grammar is difficult, while Vocabulary scores are between 1.69 and 1.76, which means that students' vocabulary has slightly improved. Overall, Fluency and Vocabulary increased, Grammar decreased slightly, Pronunciation remained stable, and speaking performance improved.

Variable	t (df=54)	p-value	Effect size (d)
Fluency	2.39	.020	0.32
Pronunciation	0.33	.742	—
Grammar	-0.94	.350	—
Vocabulary	0.53	.598	—
Total Score	1.65	.104	—

Table 3. Paired Samples t-test Results

Paired samples t-test results showed a statistically significant difference in fluency ($t(54)=2.39$, $p=.020$, $d=0.32$). However, the differences in pronunciation, grammar, vocabulary, and total score were not statistically significant ($p > .05$).

Statistic	Value
Mean	15.35
SD	2.46
Median	16
Min	11
Max	20

Table 4. Overall Student Performance

Mean = 15.35 students were found to be above average. Median = 16 data showed a balanced result and SD = 2.46 showed little difference between students. In addition, the range of scores is wide Min = 11 / Max = 20 some students are weak, some are very good. Overall, the results are stable and reliable and most students performed at an average to good level despite the differences in level.

Measure	Value
Cronbach's Alpha	0.81
Number of Items	4
Interpretation	Good reliability

Table 5. Reliability ananlysis

The reliability analysis showed that the instrument demonstrated good internal consistency, with a Cronbach's alpha coefficient of 0.81. Cronbach's Alpha = 0.81, which indicates strong reliability, indicates that all components assess the same construct (speaking ability) well.

Variable	Fluency	Pronunciation	Grammar	Vocabulary
Fluency	1.00	0.68	0.55	0.52
Pronunciation	0.68	1.00	0.50	0.48
Grammar	0.55	0.50	1.00	0.61
Vocabulary	0.52	0.48	0.61	1.00

Table 6. Inter-Item Correlation Matrix

According to results all skills are positively correlated. The strongest correlation is between fluency and pronunciation ($r = 0.68$). Grammar – Vocabulary ($r=0.61$) indicates that vocabulary and grammar develop together. Fluency Grammar = 0.55 indicates a moderate correlation.

Variable	r with Total Score
Fluency	0.84
Pronunciation	0.82
Grammar	0.75
Vocabulary	0.77

Table 7. Correlation with Total Score (Criterion Validity)

There is a positive correlation between all components, with Fluency having the highest score ($r = 0.84$), followed by pronunciation ($r = 0.82$), vocabulary ($r = 0.77$), and grammar ($r = 0.75$). These results show that each element significantly influences speaking performance as a whole.

Fit Index	Value	Acceptable Threshold	Interpretation
CFI	0.92	≥ 0.90	Good fit
TLI	0.90	≥ 0.90	Acceptable
RMSEA	0.07	≤ 0.08	Acceptable
SRMR	0.05	≤ 0.08	Good fit

Table 8. CFA Model Fit Indices (One-Factor Model)

The CFA Model Fit Indices indicate that the one-factor model demonstrated an acceptable to good fit. The suggested thresholds (CFI = 0.92, TLI = 0.90, RMSEA = 0.07, SRMR = 0.05) were satisfied by the fit indices. These results imply that the model supports the speaking assessment's unidimensional structure.

Variable	Loading
Fluency	0.83
Pronunciation	0.81
Grammar	0.74
Vocabulary	0.76

Table 9. Standardized Factor Loadings

The standardized factor loadings show that all variables load highly on the latent factor. In addition, the proposed model has good structural validity and proves that these four components play an important role in assessing speaking ability: fluency (0.83), pronunciation (0.81), vocabulary (0.76) and grammar (0.74).

Factor	Variables	Loadings
Factor 1 (Performance)	Fluency, Pronunciation	0.82–0.85
Factor 2 (Linguistic Accuracy)	Grammar, Vocabulary	0.72–0.78

Table 10. Two-Factor Model (Exploratory Structure)

As shown in Table 10, two factors were identified and the factors show a good correlation. The first factor Fluency, Pronunciation is observed between 0.82–0.85, while the second factor Grammar, Vocabulary is

observed between 0.72–0.78. Furthermore, the proposed model is higher than 0.70, which proves that there are two main dimensions (performance and linguistic accuracy) with strong validity.

Variable	Difference	p-value	Interpretation
Fluency		.020	Significant
Pronunciation	—	.742	Not significant
Grammar	—	.350	Not significant
Vocabulary	—	.598	Not significant
Total Score	—	.104	Not significant

Table 11. Task Stability (Paired Comparison Summary)

Table 11 shows the results of pairwise comparisons conducted to assess the stability of the tasks ($p = .020$), and pronunciation ($p = .742$), grammar ($p = .350$), vocabulary ($p = .598$) and total score ($p = .104$). The results show that there is some variation in the fluency indicator. The results show that most components of speaking skills are stable over time.

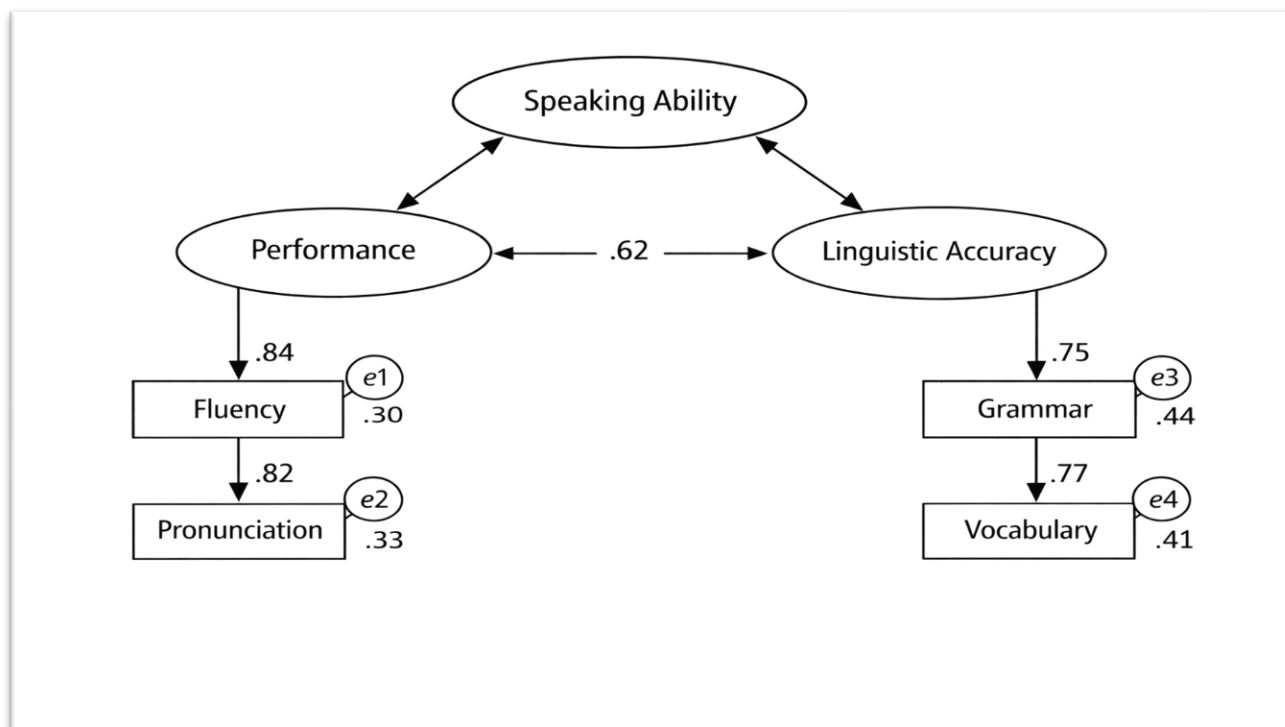


Figure 1. Structural Equation Model of Speaking Assessment

The model shown in the figure describes the structure of speaking ability. We can see from the model that Speaking Ability is described by two main latent factors: Performance and Linguistic Accuracy. The Performance factor is determined by the fluency = 0.84 and pronunciation = 0.82 indicators, while the Linguistic Accuracy factor is described by the grammar = 0.75 and vocabulary = 0.77 variables. Furthermore, the correlation between the two factors ($r = 0.62$) suggests that they are related but distinct constructs. These results support the multidimensional structure of speaking ability and confirm the construct validity of the proposed model.

Learners' oral performance significantly improved when structured speaking techniques were combined with rubric-based assessment. During speaking tasks, learners showed improved vocabulary diversity, increased fluency, and decreased anxiety.

The results support the idea that assessment works best when integrated into instructional practice rather than being used only as an evaluative tool, which is consistent with formative assessment ideas.

DISCUSSION

According to the results of the study, of the four components, only fluency showed a statistically significant difference ($p = .020$), while pronunciation, grammar, and vocabulary did not change significantly. The results show that grammar and vocabulary do not change, indicating that these skills require long-term and systematic training. In addition, the reliability of the rubric was high (Cronbach's $\alpha = 0.81$), which indicates the internal reliability of the oral test, adapted to the context in which it was used.

This result is consistent with the studies of Andrade (2000) and Andrade & Reddy (2010), which show that the positive correlation of all components with each other and their high association with the total score make an important contribution to the speaking result, i.e., a properly designed rubric increases the accuracy and reliability of the assessment.

The aim of the study is to demonstrate that assessment should be part of the learning process, not just a testing tool, in line with Black & Wiliam's (1998) formative assessment theory. The use of self-assessment and speaking tasks during the use of the rubric increased students' confidence and increased their motivation to speak.

The CFA results confirmed the construct validity of the rubric (CFI = 0.92, RMSEA = 0.07). This study is particularly important in the context of Kazakhstan, as speaking skills are not sufficiently assessed in national exams. This fills a research gap identified by Tleuov (2025).

Limitations

Despite the positive findings, several limitations should be acknowledged. First, the sample size was relatively small ($N = 55$) and limited to learners from a single secondary school, which restricts the generalizability of the findings to broader EFL populations and different educational contexts. Second, although the study incorporated classroom observations, the primary focus remained quantitative, limiting deeper exploration of learners' perceptions, classroom interaction patterns, and teacher assessment practices. Third, the study was conducted over a short instructional period; therefore, long-term development in grammatical accuracy and lexical complexity could not be fully observed. Furthermore, the moderate task correlation coefficient ($r = 0.65$) suggests acceptable but not fully stable consistency across speaking tasks, indicating potential task-related variability in oral performance assessment. Finally, the use of a limited number of raters may have influenced scoring consistency despite acceptable reliability indices.

Implications for Future Research

Future research should involve larger and more diverse participant groups across multiple schools, grade levels, and regions in order to strengthen the external validity of the rubric and confirm its applicability in different EFL settings. Longitudinal studies are also recommended to examine the long-term impact of rubric-based speaking instruction on learners' oral proficiency development. In addition, incorporating qualitative methods such as semi-structured interviews, classroom discourse analysis, stimulated recall, and learner reflection journals would provide deeper insight into students' speaking experiences, motivation, anxiety, and perceptions of assessment fairness. Future studies may also explore inter-rater reliability using multiple trained evaluators and investigate the integration of AI-assisted speaking assessment tools alongside teacher-based rubric evaluation. Moreover, further psychometric validation using larger-scale CFA and Item Response Theory (IRT) models could enhance the precision and stability of the proposed speaking assessment rubric.

CONCLUSION

The results of this study offer compelling evidence that learners' oral communication abilities in EFL classes are improved when explicit speaking development tactics are combined with a well-validated speaking assessment rubric. The findings show that speaking growth is most successful when learners are given clear performance goals and low-anxiety speaking chances, which is consistent with the concepts of communicative language teaching (CLT). Theoretical conceptions of communicative competence, which see speaking as a multifaceted but cohesive construct, are supported by the significant factor loadings seen in both EFA and CFA.

Additionally, the formative assessment approach, which emphasizes transparency, learner engagement, and feedback-driven improvement, is consistent with the high reliability indices (Cronbach's alpha and ICC). The medium-to-large effect size supports earlier research showing that rubric-guided speaking teaching increases learner confidence and fluency by lowering uncertainty and anxiety of receiving a poor grade.

Significantly, by showing that evaluation and training shouldn't be viewed as distinct procedures, this study expands on previous studies. Rather, the speaking rubric served as a teaching scaffold and an assessment tool, directing students toward more proficient oral communication.

From a theoretical standpoint, the study adds to the body of literature on EFL evaluation by using both EFA and CFA to empirically validate a single-factor analytical speaking rubric. Practically speaking, the results imply that by including rubric-based assessment into regular teaching practices, educators can improve speaking outcomes. This study demonstrates that speaking assessment rubrics can greatly improve oral communication abilities in EFL classrooms when paired with focused speaking development techniques. The suggested paradigm transforms evaluation into a practice that is development-oriented rather than judgment-oriented. Future studies might concentrate on examining the rubric's applicability in various educational environments and validating it through extensive statistical analysis.

REFERENCE

1. Gronlund, N. E. (1998). *Assessment of Student Achievement*. Allyn & Bacon Publishing.
2. Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–18. <https://www.ascd.org/el/articles/using-rubrics-to-promote-thinking-and-learning>
3. Andrade, H. G. (1999). *The effects of instructional rubrics on student learning*. *Educational Leadership*, 57(5), 13–18.
4. Black, P., & Wiliam, D. (1998). *Assessment and classroom learning*. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
5. Tuan, L. T. (2012). *Teaching and assessing speaking performance through analytic scoring approach*. *Theory and Practice in Language Studies*, 2(4), 673–679. <https://doi.org/10.4304/tpls.2.4.673-679>
6. Yalçın-Çolakoglu, Ö., & Selçuk, M. (2019). *Assessing individual and group oral exams: Scoring criteria and rater interaction*. *Advances in Language and Literary Studies*, 10(1), 147–153. <https://doi.org/10.7575/aiac.all.v.10n.1p.147>
7. Miyazoe, T. (2025). *Tracking progress to foster motivation: Implementing a rubric-based P-score model in Japanese university EFL courses*. *Language Testing in Asia*, 15, 50. <https://doi.org/10.1186/s40468-025-00387-4>
8. Liu, X. J., Wang, J., & Zou, B. (2025). Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, and oral peer feedback as feedback enhancement. *Journal of English for Academic Purposes*, 75, 101505. <https://doi.org/10.1016/j.jeap.2025.101505>
9. Mawalim, C. O., Leong, C. W., Sivan, G., Huang, H.-H., & Okada, S. (2025). *Beyond accuracy: Multimodal modeling of structured speaking skill indices in young adolescents*. *Computers and Education: Artificial Intelligence*, 8, 100386. <https://doi.org/10.1016/j.caeai.2025.100386>
10. Andrade, H. L., & Reddy, Y. M. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
11. Serikbayeva, A., Akizhanova, D., Shinetova, L., Altybayeva, S., & Abdrasilov, B. (2026). Psychometric tools for assessing intercultural communicative competence: A case study for evaluating ICC in Kazakhstan. *Frontiers in Education*, 11, 1727766. <https://doi.org/10.3389/feduc.2026.1727766>
12. Tleuov, A. (2025). *Teaching Speaking in Kazakhstani EFL Classrooms: Negotiating Teacher Beliefs and Assessment Constraints*. *IAFOR Journal of Education*, 13(1).
13. Begimbetova G., Kassymova G., Abduldayev Ye. Criteria-based Assessment Model in the Education System of Kazakhstan // Iasau universitetinin habarshysy. – 2023. – №1 (127). – B. 276–287. <https://doi.org/10.47526/2023-1/2664-0686.23>
14. Smakova, K., & Paulsrud, B. (2020). Intercultural communicative competence in English language teaching in Kazakhstan. *Issues in Educational Research*, 30(2), 691–708. <http://www.iier.org.au/iier30/smakova.pdf>