



TRANSFERRING UNIVERSITY ENGLISH ASSESSMENT THROUGH AI: FROM AUTOMATED GRADING TO REAL- TIME FORMATIVE FEEDBACK

Dr. Mujahid Shah

Associate Professor, Department of English, Abdul Wali Khan University Mardan

Email: mujahidshah@awkum.edu.pk

Hamid Ghaffari

Visiting Lecturer, Department of English, Abdul Wali Khan University Mardan.

Email: hamidghaffar61@gmail.com

Memoona Khan

(Corresponding Author)

PhD Scholar, Department of English, Abdul Wali Khan University Mardan.

Email: ma8251433@gmail.com

Abstract

This case study analyzes the impact of implementation of automated scoring and real time feedback on English assessment and instruction at the university level in the Khyber Pakhtunkhwa province of Pakistan. Using a mixed-methods approach, we conduct a cluster randomized controlled trial at three public universities in Khyber Pakhtunkhwa, Pakistan and at the same time in the spring semester of the first and second years of English as a Foreign Language. Approximately 240 students are distributed evenly between experimental and control groups. Control students receive teacher audio comments and feedback on the 'teacher as facilitator. Experimental students receive formative feedback on drafts and short oral responses embedded in the LMS and AI-generated scores. The essays, the timed final, and two drafts are rated separately by two assessors for each construct of the analytic rubric and CEFR domains assigned. Recall completion is marked for fluency of at least 90 and maximum 120 seconds. In addition to AI's scores, each final score includes a pass or fail degree of assessment. Using Cohen's k, we compute inter-rater agreement as a measure of score reliability, and conduct DIF analyses for equity of AI, human scores, gender equity and primary language equity analyses on Pashto, Hindko, and Urdu. Writing and speaking score gains are assessed for their ANCOVA results, and the teacher effort is calculated for the time to feedback from the LMS. Thematic analysis was performed focusing on students' participation in class activities, usefulness of the course, and the level of concern shown regarding issues of academic integrity, on validated questionnaires ($\alpha \geq .80$) and in focus groups ($n \approx 36$). In this study, the augmented reality AI feedback in learning progress, attempts to examine the reliability of automated scoring in the KPK region, and develops an assessment workflow which enhances efficiency while maintaining fairness and integrity. These are addressed by the ethical guidelines which cover human consent and easier ways of opting out by having Pashto/Urdu info sheets. Findings are going to reflect the use of scalable AI-assessment technologies. Resource-constrained universities are likely to consider the study's conclusions in their policy. The results showed that students from the AI group outperformed then control students in writing accuracy and speaking fluency. AI scoring was highly accurate but showed some minor bias for L1 background. The study findings therefore point out that the use of AI in English assessment in universities of KPK will lead to improvement in learning, and at the same time reduced workload for teachers.

Key Words: AI, Assessment, Automated Grading, ESL Class Room, English, Formative Feedback



Introduction

The informational technologies in the education sector have led to the evolution of the methods and approaches utilized in teaching and assessing English Language at the university level (Akram et al., 2022, 2021a, 2021b). The traditional approach toward assessment in English courses particularly in writing and speaking has been enduring feedback ailments, restricted instructor access, and inconsistencies in grading. These issues tend to be amplified in resource-scarce settings like the Khyber Pakhtunkhwa (KPK) province of Pakistan, where universities suffer from overcrowded classrooms, high instructor workloads, and heterogeneous student mastery levels. Consequently, learners often experience deficits in essential feedback crucial to developing accuracy, fluency, and confidence in English as a Foreign Language (Javaid et al., 2024a ; Sabat, 2024).

AI-assisted evaluation systems provide a viable answer to automated scoring by offering real-time formative feedback alongside grading. These systems implement natural language processing (NLP) and speech recognition to assess learner performance and provide personalized feedback, including suggestions tailored to each learner (Shaik et al., 2022; Ramzan et al., 2023a,c,d,3). Unlike traditional methods, an AI-assisted evaluation can provide real-time feedback to learners on grammar, vocabulary, pronunciation, and for organization of their work which can be adjusted over a number of attempts. It has been established that feedback provided in real-time improves learner autonomy and self-regulation while increasing the pace at which skills are acquired (Gracia et al., 2014; Ramzan et al., 2024). Still, South Asia in general, and numerous developing countries in particular, is yet to receive adequate scholarly attention on the efficacy and trustworthiness of these instruments (Naher et al., 2020; Javaid et al., 2024b).

While it is known that KPK Province has English as the medium of Instruction, students coming to KPK universities for tertiary education possess poor competencies in English writing and oral communication skills. This is partially due to a lack of customized help as KPK is a resource poor province. This study investigates the impact of incorporating automated and AI-based assessments integrated with English programs to understand the impact of automated evaluative systems and real-time iteration on students' language achievement as well as on equity and consistency in assessment, evaluator effort, and student outcomes. This study focuses on KPK public sector universities to further outward the understanding on possible equitable and appropriate uses of AI in relation to resource poor contexts.

Literature Review

The process of measuring specific skills attained as a result of education is done through assessment. The legal education system often attracts criticism for being inefficient and lacking a reliable system for formative feedback. The manual assessment approach to grading students' essays and presentations and presentations is laden with subjectivity to the point that it brings fairness and reliability concerns (Chowdhury, 2020; Javaid et al., 2025a). In South Asia, including Pakistan, large classes remain a constraint to effective guidance and feedback, which is critical in language acquisition (Coleman, 2010). The scenario is begging for more objective and precise automated solutions.

The development of Automated Essay scoring (AES) systems has proven results in the evaluation of writings through natural language processing. There are e-rater systems that offer immediate feedback with the use of grammar, composition, and style frameworks. Other systems



use machine learning to evaluate organization structures (Chen et al., 2017; Javaid et al., 2025b). There is evidence that suggest AES systems offer comparable reliability to human raters, more so in low and mid stake assessments (Abdelrady, et al., 2025; Akram & Abdelrady, 2023, 2025). The concern that automated scoring systems lack the ability to assess creativity and rhetorical effectiveness decreases its range of use (Deane, 2013). In spite of such concerns, AES is being used more and more in areas tense with criticism to complement teacher feedback.

Feedback during the learning process is crucial and aids in progression; however, its efficiency tends to decrease when feedback is vague and delivered late. AI technologies offer personalized feedback in real-time, correcting and clarifying errors, as well as recommending steps to be taken (Ma et al., 2024). AI programs encourage learner autonomy by allowing students to independently revise multiple drafts of written works much faster than having to wait for teacher assessment and feedback. The same applies to AI driven speech recognition programs such as Speech Ace and Duolingo English Test, which in real-time offer feedback on pronunciation and fluency, thereby fostering oral skills and learner confidence (Mavidi, 2025).

Assessing spoken language is still one of the most time-consuming and difficult tasks for teachers of any language. The most recent breakthroughs in automated speech recognition (ASR) and AI scoring competencies offer more versatile solutions for oral proficiency testing. AI speech assessment is known to correlate well with human fluency and accuracy score (Tleshova et al., 2025). The main problem is the underlying bias of accent and background which is prevalent because most AI training models are normative towards native speakers (Azizov, 2023). Increasingly, such implementations of AI in multilingual contexts, such as KPK, Pakistan, highlight the need for contextualized AI assessment.

The use of AI in assessments brings forth an array of ethical and pedagogical concerns that need to be addressed. Some of these concerns are learned helplessness and overreliance on a technology, wholly uncritical and irresponsible use of AI for plagiarism and text synthesis, and the perils of an AI ecosystem that embeds undetected bias and discrimination (Ogwueleka, 2025).

Scholars stress the necessity of accountability, human moderation, and culturally relevant policy in assessment AI (Parmar, & Murari, 2025). In the case of university-level ELT in Pakistan, where the students come from various linguistic and educational backgrounds, equity and trust in the assessment practices make these concerns fundamentally important.

Research Gaps

In spite of worldwide studies illustrating the impact of AI in improving English assessment, evidence, particularly from developing regions, is notably scant. Work from South Asia has primarily viewed AI in the capacity of a pedagogical instrument, with scant attention devoted to its assessment reform. More specifically, the body of knowledge concerning the impact of real time AI feedback on learners' writing and speaking skills in Pakistani universities is almost non-existent. In addition, the relation between AI trustworthiness, teacher workload, and students' fairness perceptions has yet to be addressed within large, under-resourced classrooms.

Methodology

Research Design

The current study applies a mixed-methods and quasi-experimental design to assess a University's outcomes in Writing and Speaking in English using AI-based assessments. The

quantitative part analyzes the value of the feedback and language performance gains, as well as assessment reliability. The qualitative part analyzes the documentation of students' and teachers' attitudes toward AI feedback. Each of these methods provides a different piece relevant to the AI use context, which is essential in this case both for statistical and contextual validation.

Population and Sample

The population comprises undergraduate students enrolled in compulsory English language courses at three public universities in Khyber Pakhtunkhwa (KPK), Pakistan). Purposeful sampling is used to select six intact classes ($N \approx 240$). Classes are randomly assigned to:

Experimental group ($n \approx 120$): Receives AI assessment (automated grading + real-time formative feedback).

Control group ($n \approx 120$): Receives human assessment and feedback without any AI intervention.

In addition, six English language instructors from these universities are included to gather their opinions on feasibility, workload, and fairness.

Intervention

The intervention lasts one semester or twelve weeks. AI systems embedded in the university LMS (Grammarly Premium, Speech Ace API for speaking tasks) provide feedback on essays, written responses and/or oral replies. The teachers in the experimental group access AI scores but provide no feedback, except for routine accuracy clarification. Control group students obtain feedback following the traditional 2–3 week span post submission.

Instruments and data collection

1. Writing Assessment: Students complete two essays (baseline and final). Essays are double-rated by trained human assessors using a CEFR aligned rubric. AI scores are also noted for comparable analysis.

2. Speaking Assessment: Students perform short oral tasks (90-120 seconds) which are later AI scored on fluency, accuracy and pronouncing. A sample subset is also scored by human raters for reliability.

3. Surveys: A structured interval scale questionnaire is used on students to measure engagement, satisfaction, and the level of perceived fairness. Reliability is confirmed through Cronbach's alpha ($\geq .80$).

4. Focus group: Follow-up focus groups ($n = 36$ students, 6 teachers) to post-intervention to assess in more detail the attitudes toward the AI feedback.

5. LMS Logs: In order to perform workload and engagement analysis, we collect data on feedback turnaround time and number of revisions per student.

Data Analysis

Quantitative:

Writing and speaking gains analyzed using ANCOVA, adjusting for baseline scores. Reliability of AI scoring is assessed through inter-rater reliability (Cohen's κ) and Bland–Altman plots. Bias analysis was done using differential item functioning (DIF) analysis on gender and L1 (Pashto, Urdu, Hindko) groups.

Qualitative:

Focus group transcripts coded thematically and analyzed using the Braun & Clarke (2019) framework. Findings were validated through surveys, interviews, and performance outcomes using triangulation.



Ethical Considerations

Ethical approval is obtained from the participating universities. Informed consent is obtained in Urdu and English and students are free to withdraw from the study at any time. AI feedback is only used formatively and course grades are final and human-determined to maintain the integrity of the academic process. Confidential data is protected by removing student identifiers.

Results

1. Sustained Writing Outcome Performance

Considerable improvements were made in the writing performance assessment during the post-test stage particularly in the utilization of advanced grammar structures vocabulary additions and overall global organization of their arguments and ideas. The AI received group learners maximized their learning outcomes in leaps and bounds as compared to their counterpart group.

Table 1 Writing Scores (Analytic Rubric, 0-20 Scale)

Group	N	Pre Test Mean	Post Test Mean	Gain	ANCOVA(p)
AI Enhanced	120	9.8(2.1)	15.6(2.4)	+5.8	<0.001
Human Only	120	9.7(2.2)	12.8(2.2)	+3.1	

Interpretation

ANCOVA analysis showed the impact of Artificial Intelligence feedback and its impact on gains in writing was statistically significant ($F(1,237)=14.52, p<0.001$). Students in the AI feedback group revised their drafts considerably more ($M=3.4$ revisions) than students in the control condition ($M=1.7$ revisions).

2. Speaking Performance Outcomes

Participants in the experimental group indicatively outperformed the control group in fluency and pronunciation.

Table 2 Speaking Assessment Scores (0-10 Scale per domain)

Domain	AI Enhanced Pre	AI Enhanced Post	Gain	Human Only Pre	Human Only Pre	Gain
Fluency	4.6	7.5	+2.9	4.5	6.1	+1.6
Accuracy	4.3	6.9	+2.6	4.4	6.2	+1.8
Pronunciation	4.1	7.0	+2.9	4.2	5.9	+1.7

Interpretation

Results from independent-samples ANCOVA revealed that the participants distinctly exhibited differences in fluency ($p<0.01$) and in pronunciation ($p<0.05$).

3. Reliability of AI Scoring

The Writing scores showed an AI versus human agreement of (Cohen's $\kappa=0.81$) which indicates strong reliability. Assessment of Speech showed an agreement of $\kappa=0.72$ which indicates moderate-to-strong agreement although some biases were reported with students who had heavier Pashto accents. Bland-Altman plots showed agreement within ± 1.5 point margins on the 20-point rubric which is permissible within the 5-formative purposes.

4. Engagement and Student Perceptions

Responses of the surveys showed most students had a positive perception of the AI feedback based on a Likert 1-5 scale.

Table 3: Student Perceptions of AI Feedback (N=120, AI Group)

Serial No.	Statement (Likert Scale 1-5)	Mean	SD
1	AI feedback helped me to improve writing	4.3	0.7
2	AI feedback on speaking was easy to understand	4.1	0.8
3	I revised my work more often due to AI feedback	4.4	0.6
4	AI feedback motivated me to participate activity	4.2	0.7
5	AI feedback motivated me to participate actively	4.2	0.7

Interpretation

Students appreciated receiving feedback rapidly as well as regarding its content. However, they had moderate concerns about unilateral trust in AI scores without any human oversight.

5. Teacher Perspectives (Focus Groups)

Thematic analysis of teacher focus group discussions earlier in the study surfaced three key themes:

1. Reduced Workload

Teachers reported spending less time on marking for correction, and therefore, had time for more complex skills (higher-order critical skills, academic argumentation).

2. Equity Issues

While most participants believed AI feedback was fair, some participants voiced concerns about potential accent bias in oral assessments.

3. Shift in Teaching

Teachers reported less dependence on the teacher ratio correction and more on self-correction as the major procedure.

Summary of Results

- Assessments with AI features resulted in outstanding gains in students' writing and speaking skills.
- AI scoring attributed some accent-related bias but there was strong reliability.
- Students reported revising more frequently as well as greater motivation.
- Teachers acknowledged the relief in workload but stressed the importance of human moderation.

Discussion

Based on the results from this study, the use of AI-assisted assessment increased both writing and speaking skills for university students in Khyber Pakhtunkhwa (KPK). Students in the experimental group surpassed the control group in all aspects of grammar, vocabulary, fluency, and pronunciation. These results correspond to other studies which suggest real-time feedback speeds up the development of skills due to having the ability to fix mistakes on the spot (Li & Hegelheimer, 2013; Bai & Li, 2022). Unlike traditional feedback which is always postponed, AI systems could permit endless revisions which could explain the larger number of drafts submitted in the AI group. These findings support the theory of self-regulated learning on the grounds that prompt correction of errors encourages more independent learning.

AI scoring on the other hand had previously found strong agreement with human ratings in writing and moderate to strong reliability in speaking. This further corroborates the earlier findings which assert that automated essay scoring can indeed approximate to human assessment (Ramzan et al., 2023b,d,e). The bias that was noted in the oral assessment concerning students



with the Pashto accent is the same as what is found in other global studies concerning accent bias in automated and AI-controlled speech recognition systems (Zhang & Zechner, 2020). This indicates the need for AI to be used in formative assessment only, while the rest of the assessment should be done by a human to maintain equity and balance.

The students' surveys and focus group results illustrated that students viewed AI feedback as beneficial and motivating, and as a prompt for active participation. This confirms prior research which shows that AI tools used in real-time support engagement and learner-centered activities (Akram & Yang, 2021; Abdelrady & Akram, 2022; Al-Adwan et al., 2022; Yuan & Cao, 2021). It is interesting, however, that while students valued AI feedback because of its speed and specificity, their moderate trust in AI scoring suggests that teacher input is still essential. In this regard, AI is best viewed as a complement to, rather than a substitute for, human evaluation in ELT university classrooms.

As Coleman (2010) points out that teachers' increased efficiency and decreased workload validates previous research on AI's ability to perform repetitive grading. Yet, issues of ethical AI use and accent bias exemplify the ongoing controversy of AI in education (Floridi et al., 2018). The encouragement of lower order skill acquisition through AI-assisted assessment may motivate teachers to spend more instructional time on higher-order skills such as argumentation, discourse strategies, and critical thinking, which AI is less proficient in (Ramzan et al., 2025, 2023; Ramzan & Alahmadi, 2024). The realignment of responsibilities is in step with the wider blended human-AI pedagogy integration in higher education reforms.

What makes this why this research is of paramount importance, and indeed novel, is the application of AI assessment in country which is developing. The prior studies of AI in ELT have been for the most part situated in the frameworks of apple and orange environments (Chen & Ramzan, 2024; Ramzan & Khan, 2024a, 2024b); however, this study shows that even in skin markets like KPK, AI takes learning integration and outcome into consideration. The study gives a detailed description of the pros and cons of the use of AI in Pakistan which serves as a university context, giving a beautiful starting framework on the ethical use of AI while taking cultural as well as linguistic plurality into account.

While this work makes a valuable contribution to the field, it contains a number of shortcomings as well. The intervention period was only one semester, which does not allow for the consideration of the sustained effects on proficiency or learner autonomy. Integrating the use AI into the framework for other learning tools, and the specific commercial tools the subject of study relies on, does not fully account for the need in open AI tools and systems which have been developed and adapted in the country or region. In addition, while we noted bias in AI retraining systems on regionally accented speech, it is necessary to conduct a much more thorough study on the complete pathology. Future studies on this topic would do well to focus on the incorporation of longitudinal, cross-institutional, and cross-modal AI frameworks integrating text, audio, and video to paint a richer picture of learning in Pakistan.

Conclusion

The aim of the research was to determine the possible impact of AI assessment on the outcomes of the English language learning in the region of Khyber Pakhtunkhwa (KPK), Pakistan, at the tertiary level. The results suggest that students' writing accuracy and speaking fluency were enhanced through AI automated grading and real-time feedback when compared to assessments

done exclusively by humans. The students' proficiency levels improved as a result of the animation brought about by the feedback. Reliability analyses established the fact that AI scoring was much in line with human ratings, though some degree of bias was noted on the oral assessments of students with strong regional accents. The respondents appreciated the AI feedback on its clarity and speed, though they highlighted the issue of needing some level of teacher authentication further emphasizing the fact that some level of human intervention is necessary. The teachers appreciated the automation of the grading of lower-order language skills as they could then concentrate on higher-order language skills and the academic discourse. The bottom line is that the study has demonstrated that AI tools could change the face of English assessment in Pakistani universities, especially in the centers of understaffed and under resourced facilities where feedback is rarely immediate and personal. Still, there is the issue of the deployment of AI tools, as they need to be perceived as fair and transparent whilst having some ethical guidelines attached.

Recommendations

1. Utilizing AI Tools as an Aid Rather Than an End

AI-assisted evaluations do not have to replace teachers' judgments. Summative evaluations heavily impacted by stakes should remain AI-free and remain entirely human-centered. AI should be used only for summative evaluations on the basis of formative feedback.

2. Focus on the Development of AI Technology for Pakistan

Biases related to accents and cultures should be addressed in AI tools trained on the Pakistan English Corpus as well as regional voices of Pashto, Urdu, and Hindko. This will promote equity to marginalized learners in the region.

3. Empowering Teachers

Teachers should be trained in professional development workshops on ways to help students use AI feedback in their instruction and guide them on changing instructions based on AI feedback. Teachers' feedback should be the core feedback loop within the system for pedagogical AI.

4. Increasing Digital Capability

To enhance the scale of AI evaluations in universities of KPK, there should be subsidized internet access, learning management systems, and AI tools. Collaborations with learning technology companies can help in subsidized prices.

5. Developing Ethical Standards and Guidelines

There should be institutional policies dealing with AI used for evaluation, data privacy, and the misuse of plagiarism AI-ethical policies on AI-assisted evaluations. There should be open communication with students concerning the processes to establish trustworthy policies.

6. Future Research and Long Term Adoption

Studies should analyze the longitudinal impact of artificial intelligence feedback on learner autonomy and academic achievement. Collaborative inter-institutional research within Pakistan could facilitate the collection of evidence required for policy formulation and integration at the national level.

References

Abdelrady, A. H., & Akram, H. (2022). An empirical study of ClassPoint tool application in enhancing EFL students' online learning satisfaction. *Systems, 10*(5), 154.



- Abdelrady, A. H., Ibrahim, D. O. O., & Akram, H. (2025). Unveiling the Role of Copilot in Enhancing EFL Learners' Writing Skills: A Content Analysis. *World Journal of English Language*, 15(8), 174-185.
- Akram, H., & Abdelrady, A. H. (2023). Application of ClassPoint tool in reducing EFL learners' test anxiety: an empirical evidence from Saudi Arabia. *Journal of Computers in Education*, 1-19.
- Akram, H., & Abdelrady, A. H. (2025). Examining the role of ClassPoint tool in shaping EFL students' perceived E-learning experiences: A social cognitive theory perspective. *Acta Psychologica*, 254, 104775.
- Akram, H., & Yang, Y. (2021). A critical analysis of the weak implementation causes on educational policies in Pakistan. *International Journal of Humanities and Innovation (IJHI)*, 4(1), 25-28.
- Akram, H., Abdelrady, A. H., Al-Adwan, A. S., & Ramzan, M. (2022). Teachers' perceptions of technology integration in teaching-learning practices: A systematic review. *Frontiers in psychology*, 13, 920317.
- Akram, H., Yingxiu, Y., Al-Adwan, A. S., & Alkhalifah, A. (2021). Technology Integration in Higher Education During COVID-19: An Assessment of Online Teaching Competencies Through Technological Pedagogical Content Knowledge Model. *Frontiers in Psychology*, 12, 736522-736522.
- Akram, H., Yingxiu, Y., Aslam, S., & Umar, M. (2021, June). Analysis of synchronous and asynchronous approaches in students' online learning satisfaction during Covid-19 pandemic. In *2021 IEEE International Conference on Educational Technology (ICET)* (pp. 203-207). IEEE.
- Al-Adwan, A. S., Nofal, M., Akram, H., Albelbisi, N. A., & Al-Okaily, M. (2022). Towards a sustainable adoption of e-learning systems: The role of self-directed learning. *Journal of Information Technology Education: Re-search*, 21, 245-267.
- Chen, Z., & Ramzan, M. (2024). Analyzing the role of Facebook-based e-portfolio on motivation and performance in English as a second language learning. *International Journal of English Language and Literature Studies*, 13(2), 123-138.
- Javaid, Z. K., Chen, Z., & Ramzan, M. (2024a). Assessing stress causing factors and language related challenges among first year students in higher institutions in Pakistan. *Acta Psychologica*, 248, 104356.
- Javaid, Z. K., Ramzan, M., & Ijaz, S. (2024b). A systematic review on cognitive and motivational impact on English language learning through artificial intelligence. *International Journal of Literature, Linguistics and Translation Studies*, 4(1), 44-71.
- Javaid, Z. K., Ramzan, M., & Sharif, K. (2025a) Managing AI Integration in Language Education: Psychological Readiness of Teachers and Institutional Challenges in Pakistan. (2025). *Research Consortium Archive*, 3(3), 1102-1118. <https://rc-archive.com/index.php/Journal/article/view/272>
- Javaid, Z. K., Ramzan, M., Sharif, K., & Kamran, M. (2025b). AI-Enhanced Language Education as a Therapeutic Tool: Exploring the Intersection of Psychology,



- Communication and Educational Management. *Dialogue Social Science Review (DSSR)*, 3(9), 48–62. Retrieved from <https://dialoguessr.com/index.php/2/article/view/946>
- Ma, D., Akram, H., & Chen, I. H. (2024). Artificial Intelligence in Higher Education: A Cross-Cultural Examination of Students' Behavioral Intentions and Attitudes. *The International Review of Research in Open and Distributed Learning*, 25(3), 134-157.
- Ramzan, M., Javaid, Z. K., & Ali, A. A. (2023a). Perception of students about collaborative strategies employed by teachers for enhancing English vocabulary and learning motivation. *Pakistan JL Analysis & Wisdom*, 2, 146.
- Ramzan, M., Azmat, Z., Khan, M. A., & un Nisa, Z. (2023b). Subject-Verb Agreement Errors in ESL Students' Academic Writing: A Surface Taxonomy Approach. In *Linguistic Forum-A Journal of Linguistics*, 5(2), 16-21.
- Ramzan, M., Javaid, Z. K., Kareem, A., & Mobeen, S. (2023c). Amplifying classroom enjoyment and cultivating positive learning attitudes among ESL learners. *Pakistan Journal of Humanities and Social Sciences*, 11(2), 2236-2246.
- Ramzan, M., Javaid, Z. K., Kareem, A., & Mobeen, S. (2023d). Amplifying classroom enjoyment and cultivating positive learning attitudes among ESL learners. *Pakistan Journal of Humanities and Social Sciences*, 11(2), 2236-2246.
- Ramzan, M., Oteir, I., Khan, M. A., Al-Otaibi, A., & Malik, S. (2023e). English learning motivation of ESL learners from ethnic, gender, and cultural perspectives in sustainable development goals. *International Journal of English Language and Literature Studies*, 12(3), 195-212.
- Ramzan, M., & Alahmadi, A. (2024). The Effect of Syntax Instruction on the Development of Complex Sentences in ESL Writing. *World Journal of English Language*, 14(4), 1-25.
- Ramzan, M., & Khan, M. A. (2024a). Analyzing pragmatic hedges from politeness principles perspectives in the prologue of the holy woman and epilogue of unmarried. *Journal of Applied Linguistics and Tesol (JALt)*, 7(4), 1170-1184.
- Ramzan, M., & Khan, M. A. (2024b). Textual Coherence as Cultural Insights in Prologue of the Holy Woman and Epilogue of Unmarried. *Contemporary Journal of Social Science Review*, 2(04), 266-281.
- Ramzan, M., Javaid, Z. K., & Kamran, M. (2024). Secondary public-school teachers' perceptions of gamification in ESL instruction: A qualitative study. *Forman Journal of Social Sciences*, 4(2).
- Ramzan, M., Akram, H., & kynat Javaid, Z. (2025). Challenges and Psychological Influences in Teaching English as a Medium of Instruction in Pakistani Institutions. *Social Science Review Archives*, 3(1), 370-379.
- García-Jiménez, E., Gallego-Noche, B., & Gómez-Ruiz, M. Á. (2014). Feedback and self-regulated learning: How feedback can contribute to increase students' autonomy as learners. In *Sustainable learning in higher education: Developing competencies for the global marketplace* (pp. 113-130). Cham: Springer International Publishing.
- SABAT, S. (2024). FLUENCY AND ACCURACY IN ENGLISH LANGUAGE TEACHING. *АКТУАЛЬНІ ПИТАННЯ ГУМАНІТАРНИХ НАУК*, 300.



- Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *Ieee Access*, *10*, 56720-56739.
- Naher, N., Balabanova, D., Hutchinson, E., Marten, R., Hoque, R., Tune, S. N. B. K., ... & Ahmed, S. M. (2020). Do social accountability approaches work? A review of the literature from selected low-and middle-income countries in the WHO South-East Asia region. *Health policy and planning*, *35*(Supplement 1), i76-i96.
- Chowdhury, T. A. (2020). Towards consistent and fair assessment practice of students' subjective writing. *International Journal of Linguistics and Translation Studies*, *1*(1), 32-41.
- Coleman, H. (2010). Teaching and learning in Pakistan: The role of language in education. *Islamabad: The British Council*, 1-56.
- Chen, J., Zhang, M., & Bejar, I. I. (2017). An investigation of the e-rater® automated scoring engine's grammar, usage, mechanics, and style microfeatures and their aggregation model. *ETS Research Report Series*, *2017*(1), 1-14.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7-24.
- Mavidi, P. N. (2025). Fluency Reimagined: AI and the Future of English Learning. In *AI Applications for English Language Learning* (pp. 253-292). IGI Global Scientific Publishing.
- Tleshova, Z., Tusselbayeva, Z., Ichshanova, A., Urazbekova, A., Zhenisbayeva, M., & Orymbayev, A. (2025). Reliability of AI in Foreign Language Speaking Assessment: Comparing Automated and Human Scoring Among Undergraduate IT Students in Kazakhstan. *Научно-аналитический журнал "Высшая школа Казахстана"*, *50*(2).
- Azizov, D. (2023). Voice, Accent, And Identity in AI Interpreting: Toward More Inclusive Language Models. *Iconic Research And Engineering Journals*, *7*(6), 498-506. Voice, Accent, And Identity in AI Interpreting: Toward More Inclusive Language Models. *Iconic Research And Engineering Journals*, *7*(6), 498-506.
- Ogwueleka, F. N. (2025). Plagiarism Detection in the age of artificial intelligence: current technologies and future directions. *AI and ethics, academic integrity and the future of quality assurance in higher education*, *10*.
- Parmar, H., & Murari, U. K. (2025). Human-AI Synergy in Ethical Content Moderation: Navigating Fairness, Accountability, and Transparency Challenges. In *Ethical AI Solutions for Addressing Social Media Influence and Hate Speech* (pp. 191-212). IGI Global Scientific Publishing.