

JQSS: (Journal Of Quranic and Social Studies)

ISSN: (e): 2790-5640 ISSN (p): 2790-5632

Volume: 6, Issue: 1, Jan-April 2026. P: 112-123

Open Access: <https://www.jqss.org/index.php/JQSS/article/view/229>

DOI: <https://doi.org/10.5281/zenodo.19762632>




Copyright: © The Authors Licensing: this work is licensed under a creative commons attribution



4.0 international license



## Artificial Intelligence in English Language Assessment: Opportunities and Ethical Concerns – A Balochistan-Focused Analysis

- |   |   |
|---|---|
| 1. <b>Zualfiqar Ali</b> <br><a href="mailto:Zualfiqarali9@gmail.com">Zualfiqarali9@gmail.com</a>         | PhD Scholar, Gomal University, Dera Ismail Khan, KPK, Pakistan              |
| 2. <b>Dr. Muhammad Fahim Baloch</b> <br><a href="mailto:balochfahim@gmail.com">balochfahim@gmail.com</a> | Associate Professor, Department of Media Studies, University of Balochistan |
| 3. <b>Pervez Ahmed</b> <br><a href="mailto:Pervez.uob@gmail.com">Pervez.uob@gmail.com</a>               | Assistant professor, Pakistan study center University of Balochistan        |

**How to Cite:** . Zualfiqar Ali and Dr. Muhammad Fahim Baloch and Pervez Ahmed (2026) Artificial Intelligence in English Language Assessment: Opportunities and Ethical Concerns – A Balochistan-Focused Analysis, (JQSS) Journal of Quranic and Social Studies,6(1), 112-123.

### Abstract and Indexing



### Publisher

HRA (AL-HIDAYA RESEARCH ACADEMY) (Rg)  
Balochistan Quetta





# Artificial Intelligence in English Language Assessment: Opportunities and Ethical Concerns – A Balochistan-Focused Analysis

Journal of Quranic  
and Social Studies  
112-123

**Zualfiqar Ali** 

PhD Scholar, Gomal University, Dera Ismail Khan, KPK, Pakistan

**Dr. Muhammad Fahim Baloch** 

Associate Professor, Department of Media Studies,  
University of Balochistan

**Pervez Ahmed** 

Assistant professor, Pakistan study center, University of Balochistan

© The Author (s) 2026  
Volume:6, Issue:1, 2026  
DOI:10.5281/zenodo.19762632

[www.jqss.org](http://www.jqss.org)

ISSN: E/ 2790-5640

ISSN: P/ 2790-5632

**OJS** **PKP**  
OPEN JOURNAL SYSTEMS PUBLIC KNOWLEDGE PROJECT

## Abstract

Automated Essay Scoring (AES) softwares and AI-based Speaking assessment software are being utilized in classroom and in testing Centres worldwide during integrating AI in learning and teaching. This paper outlines opportunities and ethical issues of AI integration in English language assessment, emphasizing on algorithmic bias, data privacy and overlooking role of English teacher. Based on recent empirical studies, such as those that show that there are racial and linguistic prejudices in the grading of GPT-4o essays that were so serious. Indeed, still there is a continuous discrimination due to diverse accent in automated speech recognition systems. The review has also revealed that the current AI assessment technologies, although they seem to be efficient, consistent, and scalable are fundamentally incapable of evaluating language impartially, amongst heterogeneous learners. Especially, this article pays special attention on the learners whose native languages are Balochi, Pashto, Saraiki and Punjabi limiting the scope to Pakistan generally and Balochistan particularly. This article also considers and addresses the critical infrastructural issues of Balochistan (sporadic electricity, poor internet connection, internet blackouts caused by security) and suggests a tangible, offline-first, edge-computing implementation paradigm, which can allow complex AI evaluation even in the most isolated and security-impacted districts of Balochistan. Then this article recommends a human-focused, hybrid assessment model where AI is used as a supportive tool and not as a substitute to teacher judgment that is based on transparency standards and periodic bias audits and transparent policies regarding data governance. The conclusion provides context-specific practical recommendations to educators, institutions, and policymakers who wish to use AI assessment tools in a responsible manner in the context of Balochistan during English language teaching.

**Keywords:** artificial intelligence, English language assessment, automated essay scoring, algorithmic bias, accent discrimination, Balochi, Pashto, Brahvi, Saraiki, Punjabi, data privacy, teacher oversight, offline AI, edge computing, internet shutdowns, Balochistan.

Corresponding Author Email: [zualfiqarali9@gmail.com](mailto:zualfiqarali9@gmail.com)

## Introduction

Artificial intelligence (AI) has infiltrated almost all aspects of education, yet it has perhaps found its way into language assessment with more velocity and, perhaps, with greater controversy. AI-based systems like automation of essay scoring (AES) and AI-based speaking assessment systems as well as generative AI systems like ChatGPT are slowly being used to grade student writing and speech and promise to provide feedback quickly, consistently, and at scale (Ramesh and Sanampudi, 2022; Weigle, 2013). The attractiveness of AI-assisted assessment is especially high in English as a Second Language (ESL) and English as a Foreign Language (EFL) settings, where the number of students in the classroom is frequently high, and the time of the teacher is limited. The case is particularly difficult in Balochistan, the biggest province of Pakistan in terms of the territory. The province has a very varied population of Balochi, Pashto, Brahvi, Saraiki (in border areas), and Punjabi (in urban areas). The official language of higher education and government is English, and English access to quality English education is unequal, particularly in remote districts, including Chagai, Washuk, Kharan, and Awaran (Javed and Nabi, 2022). Additionally, there are serious problems with infrastructure in Balochistan: frequent power outages (load shedding), very low broadband/internet penetration, and unreliable mobile internet in rural locations. In Turbat or Dalbandin, a teacher cannot afford to use cloud-based AI applications, which need to be constantly connected. The integration of AI-powered language tools has been a topic of recent studies conducted in Pakistan, with the results suggesting that vocabulary acquisition, grammar accuracy, and writing fluency have improved significantly, especially in urban areas with a higher level of digital access (Zahid et al., 2025). The possibility of AI-based chatbots to improve the English speaking proficiency of rural Pakistani learners has also been researched (Fatima et al., 2025). Yet, these studies have mostly concentrated on overall effectiveness, rather than on the performance of AI systems in different learners in diverse mother-tongues or the reality of infrastructure in a province such as Balochistan. This dissonance is alarming, since when an AI talking evaluation systematically penalizes an English speaker with a Balochi accent or needs an internet connection that is unavailable, it is not just grading linguistic competence- it is continuing linguistic and digital discrimination. However, these opportunities are coupled with some worrisome ethical issues which are brought into the limelight by a developing body of research. Research has reported that large language models are characterized by high levels of bias in the case of scoring essays that are written by learners of other racial and ethnic groups (Yang et al., 2024). It has been demonstrated that automated speech recognition systems are unable to handle non-native accents and penalize speakers whose voices do not match the traditional American or British accents (Subhan et al., 2025). In the case of Balochi or Pashto speaking learners, this implies that a learner who is an English speaker with a local accent might have a lower score on the AI-generated speaking score than a learner with the same linguistic competence but a closer-to-standard accent. There are also profound questions of student data privacy, transparency of algorithmic decision-making, and the role of human educators in a more automated assessment world when AI assessment

tools are used. In this article, the author looks at the opportunities as well as the ethical issues of AI in English language assessment, specifically of the learners whose native languages are Balochi, Pashto, Saraiki, Punjabi and Brahvi (the latter is not only native to Balochistan, but also unique). It then addresses the problem of infrastructure directly, such as internet freezing due to security concerns, and suggests a practical, offline-first implementation pattern, which can be functional even in areas with unstable or no power or internet connection at all. The article ends by presenting a human-focused, context-sensitive model of the responsible implementation of AI assessment tools in the diverse and challenging Balochistan environment. (Landa-Blanco, 2026; Saleem et al., 2025)

## **2. AI in English Language Assessment: Trending Tools and Capabilities.**

**2.1 Automated Essay Scoring (AES) Systems:** Automated Essay Scoring can be described as the use of a computer program to assess the quality of an essay by giving a single score, a critical assessment of the features of the essay or both (Zhang, 2021). of AES systems have been applied to the evaluation of writing for more than 20 years (Ifenthaler, 2022). Although with the development of generative AI, the systems have been able to further improve their functions. AES has a variety of benefits over human evaluation: It is time-saving (Ramesh and Sanampudi, 2022; Weigle, 2013), cost-effective (Uto and Aramaki, 2024), and saves the work of teachers (Pack et al., 2024). Modern AES systems, such as those developed based on large language models like GPT -4, are more sophisticated in analysing essays' grammar, coherence, vocabulary, argumentation and stylistic features. The tools available commercially include writing assistants like Grammarly and Trinka (Offering context-aware feedback in addition to simple grammar corrections, including academic phrase banks, citation checkers, and plagiarism detectors) and specific AES tools like Criterion and Pigai (Ding and Zou, 2024). The generative AI has also increased the opportunities: a study comparing ChatGPT and Grammarly in evaluating ESL student writing automatically detected that both systems show improved performance in writing, but some learners noted that Grammarly was more accurate and others found more detailed feedback in the ChatGPT (Ali Khan, 2024).

**2.2 AI-powered Speaking Evaluation:** Spoken English assessment is more of a technical challenge than writing assessment. Yet in recent years, automatic speech recognition (ASR) has become more advanced and allows AI-based speaking assessment to become a possibility. All of these tools are cloud-based, so they need a constant, high-speed internet connection, which is a luxury that is not possible in most of rural and peri-urban Balochistan due to its lack of stable connections. This dependency makes it unrealistic for every learners who are supposed to be most needy learners.

**3. Algorithmic Bias: A Perennial Ethical Dilemma,** In spite of their technical complexity us that AI evaluation systems cannot be seen as independent adjudicators of language proficiency. An expanding literature shows that these systems have systematic biases that disfavor some groups of learners- and in the case of Balochistani students speaking

Balochi, Pashto, Brahvi, Saraiki or Punjabi as their mother tongue, such biases are not hypothetical.

**3.1 Bias in Automated Essay Scoring:** Recent studies have revealed large-scale model bias in large language models that are employed to score essays. Yamashita (2025) carried out an experiment investigating possible biases in the ratings of GPT-4o with respect to the essays of English language learners, utilizing the ELLIPSE corpus of 6,482 essays rated by 27 human raters. The findings showed that whereas GPT-4o did not have a large bias in terms of gender or socioeconomic status, it had a strong bias in terms of race and ethnicity, with Asian/Pacific Islander group receiving higher scores and Hispanic/Latino group receiving lower scores than expected (Zheng, 2024). These findings highlight the importance of careful and critical applications of GPT as an AES tool. The same has been said regarding discrimination against English as a Second Language (ESL) learners in general. A bias study on a fine-tuned DeBERTa-v3 model on the ASAP 2.0 and ELLIPSE datasets by Fan and Yun (2026) showed that there was a limited scaling of scores on high-proficiency ESL writing, with high-proficiency ESL essays scoring 10.3 percentage points lower than native speaker essays of the same human-rated quality (Fan and Yun). The views of educators also sheds more light on these biases: a mixed-methods study revealed that educators expressed serious concerns about the meaningfulness of scores, especially the capacity of AWSF to reflect the subtleties of the writing of multilingual learners (Kim et al., 2026).

**3.2 Bias in Automated Speaking Assessment-**The Case of Balochi, Pashto, Brahvi, Saraiki, and Punjabi Speakers (with Concrete Language Examples). Bias in automated speaking assessment is even more emphasized. One of the significant issues with the traditional ASR systems is the inability to identify and evaluate non-American accents, distorting the validity of oral proficiency exams (Bandodkar et al., 2024; Zellou and Holliday, 2024). Speakers with regional or non-native accents are systematically penalised. This prejudice has tangible implications to learners who have mother-tongues of Balochi, Pashto, Brahvi, Saraiki or even Punjabi. The following are particular examples, and Brahvi has been included because it is special to Balochistan. Pashto speakers (in use in the north of Balochistan: Pishin, Qilla Abdullah, Zhob)- Pashto has no /p, /f/; so, think, now become [tink].- Dental fricatives /θ, ð/ are replaced by /t, d/; thus, think, now [tink]. Balochi speakers (central, southern, western Balochistan)- Epenthesis in initial clusters: Spring = s i p r i n g. -Devoicing final stops: Bob = b o p. -Vowel change: cat = a = e ) = k at. Brahvi speakers (mainly in Kalat, Mastung, Khuzdar, Nushki)- Brahvi has retroflex consonants (/ʈ, ɖ, ɳ, ʌ/), which are not found in English. English alveolar stops can be retroflexed: “top” [ʈap], dog [ɖɔg].- Length of the vowel is not contrastive in Brahvi; in English length contrasts are disregarded. The speakers of Saraiki (border districts such as Dera Bugti, Jaffarabad)- Implosive consonants (/ɓ, ɗ, ɟ, ɠ/) influence English stops- Transfer of the tone to English intonation. Punjabi speakers (urban centres such as Quetta)- Dental fricatives. - dental stops.- Syllable-timed rhythm. These are systematic differences, not in the communicative sense of the word, they are predictable systematic

characteristics of the L2 varieties of English. However, they are penalised by ASR systems that are trained on a native speech only, which results in unfair judgements.

**4. Balochistan Infrastructure Reality:** Internet Shutoffs and Security Restrictions on 44% of the land in Pakistan (Ahmad et al., 2024) is less than 5% of broadband connection in the entire nation. As the Pakistan Telecommunication Authority (PTA, 2024) notes, the mobile broadband penetration in rural Balochistan is lower than 15 percent, and not all of the districts are covered with 3G/4G. Solar power is widespread but not universal, and electricity load shedding is 12-16 hours a day on average in rural regions (Batool et al., 2022). In addition to these technical shortcomings, there is another, more drastic limitation to Balochistan, namely, security-related internet shutdowns. In most of the districts, especially those which border Afghanistan and Iran (e.g., Chagai, Kech, Panjgur, Washuk), mobile and broadband services are frequently disrupted by a day or even a week at a time because of security operations or civil strife (Fletcher & Hayes-Birchler, 2023; Payande & Charkameh, 2025). The government can impose a province-wide lockdown even in comparatively stable regions during sensitive times (e.g., national holidays, election days, or following militant attacks) (Sharma et al., 2023). As a result, large parts of the Balochistan region are without any internet connectivity during certain periods of the year. Real-life illustration: There is a high school in Chagai district (along the Afghan border) with 150 pupils, two English teachers and one computer which operates on a solar battery. There is no internet connection in the school. Even when the school had a patchy 2G connection, the shutdowns due to security considerations imply that that connection is not reliable or not at all during weeks. An AI talking application (e.g., Speechace) that is cloud-based is thus unusable forever. Furthermore, all systems that need to be periodically synchronized to the internet (e.g. once a week) would also not work since it cannot be ensured that the internet will be up on the set day. Therefore, an AI evaluation solution to Balochistan should be completely offline, not only on how it processes but also how it updates. It should work without any live internet connection and updates of the model should be provided through some physical medium (USB drives) that can be transported by the teachers, education officers or even through mail.

**4.1 An Offline, Edge-Computing Model with Physical Update Distribution:** We suggest a three-tier offline model that is fully autonomous with respect to access to the internet and can withstand shutdowns caused by security issues. Tier 1: Edge Device (School-Level) Each school taking part will be given a high-tolerance, low-power device (e.g., a Raspberry Pi 4 with 4GB RAM, running on a solar-charged battery). This device stores:- A lightweight, pre-trained ASR model specifically fine-tuned on a corpus of Balochi-, Pashto-, Brahvi-, Saraiki-, and Punjabi-accented English (instead of native speech).- An offline AES model (e.g., a distilled version of DeBERTa) for essay scoring.- A local database for storing student work and scores. The device creates a local Wi-Fi hotspot. Student receive poor Android tablets (they can also use their own

smartphones) to establish a connection, record oral answers, and write essays. All processing occurs to the edge device. There is no need of an internet - ever. The system is capable of operation over months or years without any outside connection.

**Tier 2: Physical Update Distribution (USB Drive Model)** Since security shutdowns do not allow a secure internet connection, model update and data export cannot rely on online synchronisation. Rather, it operates through a sneakernet mechanism:- That every academic term (e.g. every 3-4 months), a specific teacher or district education office is provided with a USB drive to which they bring the drive, and which the teacher connects to the edge device, which in turn uploads new models (and scripts to export anonymised student performance data) to the central server via a secure, government-provided connection. This circumvents the internet blockages in their entirety. Although the internet of the whole province may go offline months, the assessment system will still operate at the school level and data can be received each time a USB drive is physically received.

**Tier 3: Central Server (Quetta)** The central server consolidates the data about hundreds of schools, performs bias audits (disaggregated by mother tongue), and retrains accent-specific models. New models are transferred to USB sticks and physically handed to schools through the already established education department logistics system (e.g. textbook delivery vans, official travel of officers). Cost estimate: Raspberry Pi 4 (\$50) + solar battery (\$30) + tablet (\$60) + USB drives (negligible) = \$140 per school. In the case of 1,000 schools, 140,000, a small part of the price of cloud-based solutions and, most importantly, not depending on the internet availability.

**4.2 Concrete Pilot with Security Shutdowns in mind:** A pilot will be launched with the edge device and five tablets to each of the three districts, which represent the diversity of the province and its security problems: Pishin (Pashto majority, fairly stable, with periodic shutdowns): Khuzdar (Brahvi and Balochi mix, occasional shutdowns): Chagai (Balochi majority, frequent and prolonged Two-day training on the use of the offline interface, interpretation of AI scores, and overriding is provided to teachers. There is no need to train on internet-based tools as the system does not use the internet. The update procedure is planned in advance: a district education officer comes to each school once every term with a USB drive with updated models. Should the security conditions preclude traveling during a given term, the system still functions with the available models - no functionality reduction. The data collection will be delayed until there is safe travel. Models that are specific to the accent are fine-tuned with locally gathered speech data (100 speakers of all languages/groups) and then deployed. These models are scalded onto USB drives and installed in the course of the first set-up.

**5. Low-Connectivity, High-Security Data Privacy and Student Rights:** The offline, USB-based model actually improves data privacy since student work does not leave the school grounds other than on an encrypted USB drive that is carried in by a government official with authorisation. This is in line with the draft Personal Data Protection Bill of Pakistan and international best practices. Furthermore, since the data

does not ever pass through the internet, chances of being intercepted or unauthorised access are practically non-existent, something needed in a security-sensitive area. The teachers are still required to seek parental approval when recording the speech of students. The main server contains only de-identified metadata (no names, only anonymised codes ). USB drives are coded and logged by using a logbook.

**6. The Irreplaceable Nature of Teacher Oversight:** Arguably the most recurring conclusion of recent studies is that both teachers and students see the possible positive outcomes offered by AI assessment tools, but they do not believe in AI-only grading and demand human supervision. A mixed-method co-design pilot study with 19 K-12 educators who had an AI-based grading system discovered that although teachers found the AI-based rapid narrative feedback helpful in formative assessment, they do not trust automated scoring and believed that human intervention is necessary (Li et al., 2025). Students welcomed fast, revision-oriented feedback but remained sceptical of AI-only grading, highlighting the need for trustworthy, teacher-centred AI assessment tools (Chen, 2025). A different research involving university students and English writing teachers in China found four key themes in the perception of AI scoring systems, namely, perceived bias and inflexibility in the algorithms, the absence of transparency in the score generation process, the conflict between the authority of teachers and AI judgment, and institutional gaps in the policy and support (Kim et al., 2024). In the study, it was revealed that AI scoring systems were frequently considered pedagogically blind and ethically questionable even though they had certain operational benefits. In the education industry, an overall consensus exists that AI can only be a supplementary product and not a substitute of teachers (Aldino et al., 2026). Last grading should be left to educators especially when the assessment has direct influence to the academic progress of the students. There is a growing trend of considering a hybrid solution, which is human judgment and technology assistance as the most balanced one (Sajja et al., 2026). It is a hybrid model that maintains the best of teachers, context and creativity recognition, and empathetic feedback, but uses AI to do what it is best, large amounts of routine work processed quickly and consistently. One analysis suggested that AI grading could promote the most human aspects of instruction by providing educators with additional time to do relationship-building tasks, including more precisely customized lessons or family messages (Li et al., 2025). The Balochistan offline model involves a plain interface on the edge device to check the AI generated scores, change them or provide qualitative remarks by the teacher. The device marks suspicious cases (e.g., borderline scores or language use that is abnormal) to be reviewed by the teachers.

**7. A Human-centred Framework of responsible AI Assessment in Balochistan:** based on the above analysis, the following framework is proposed, which is specifically tailored to the Balochistan context.

**7.1 Transparency:** Schools should notify parents and students that AI is being used to assess formative assessment, how the data is stored locally, and seek permission to synchronise regularly which is every month.

**7.2 Fairness With Accent-Specific Calibration:** Rather than off-the-shelf ASR models, institutions should have models commissioned or calibrated on Balochi-accented English, Pashto-accented English, Brahvi-accented English, Saraiki-accented English, and Punjabi-accented English. To prevent systematically penalising a group, the central server should conduct quarterly bias audits, disaggregating scores by mother tongue.

**7.3 Offline -First, Solar-Powered, Edge-Computing with Physical Update Distribution:** The system should be completely offline - no reliance on any form of internet anywhere, even to receive updates. The export of model and data and subsequent model updates are done through encrypted USB drives which are transported by government officers. This architecture is resilient to the lack of infrastructure and internet outages caused by security.

**7.4 Human-in-the-Loop:** The teachers should be allowed to make the final decision in any score that influences the progression of the student. The system would indicate questionable cases to teachers.

**7.5 Data Governance:** Student data is never taken out of the school other than on encrypted USB drives. Only anonymised data is stored in the central server. The contracts with vendors (to develop the model) should not allow the secondary use of data.

## 8. Conclusion

AI has real potential of improving the assessment of the English language in Balochistan: it will be faster, more consistent, and is scalable to serve remote students who do not receive feedback at all, at present. But these opportunities can and only be achieved as long as we address the specific issues of the province lingual diversity (Balochi, Pashto, Brahvi, Saraiki, Punjabi), serious lack of infrastructure (no internet, frequent power outages), and security-related internet restrictions that render any cloud-based solution impossible. The answer is not to bring in cloud-based applications that are targeted at affluent, connected classrooms. Rather, we need to roll out offline-first, edge-computing AI, which uses solar-powered Raspberry Pi computers, trained on local accents, and supervised by teachers. Physical USB drive distribution must be used to update the model and collect data so that it is not subjected to internet shutdowns in any way. The pilot in three districts (Pishin, Khuzdar, Chagai) might be taken as an example to other parts of the province (and to other rural, security-challenged parts of Pakistan). The way ahead is not to leave AI but to responsibly adapt it: by calibrating accents, continuing bias audits, assessing human-in-the-loop, and enforcing high-quality, offline-first data governance. The professional judgment of the hardworking teachers of Balochistan

should be supplemented rather than substituting them by AI. The introduction of AI wisely will enable us to solve the issue of teacher shortage, geographical isolation, and connectivity blackouts that have historically plagued the English education in the province. If applied haphazardly, it may entrench language discrimination and online disparity. To millions of Balochistani students, at home speaking Balochi, Pashto, Brahvi, Saraiki or Punjabi, and frequently months on the internet without a single connection, the stakes could not be greater.

## References

1. Ahmad, B., Zhao, Z., Jile, X., Gultaj, H., Khan, N., & Yunxian, Y. (2024). Exploring the influence of internet technology adoption on the technical efficiency of food production: insight from wheat farmers. *Frontiers in Sustainable Food Systems*, 8. <https://doi.org/10.3389/fsufs.2024.1385935>
2. Aldino, A. A., Maheshi, B., Li, Y., Zhou, Y., Tsai, Y., Gašević, D., & Chen, G. (2026). Enhancing learner-centered feedback with AI: teachers' practices and perceptions. *Assessment & Evaluation in Higher Education*, 1. <https://doi.org/10.1080/02602938.2026.2638920>
3. Bandodkar, G., Agarwal, S., Sughosh, A. K., Singh, S., & Choi, T. (2024). "Allot?" is "A Lot!" Towards Developing More Generalized Speech Recognition System for Accessible Communication. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21), 23327. <https://doi.org/10.1609/aaai.v38i21.30381>
4. Barrot, J. S. (2025). Trink: Facilitating academic writing through an intelligent writing evaluation system. *Assessing Writing*, 65, 100953. <https://doi.org/10.1016/j.asw.2025.100953>
5. Batool, K., Zhao, Z., Atif, F., & Dilanchiev, A. (2022). Nexus Between Energy Poverty and Technological Innovations: A Pathway for Addressing Energy Sustainability. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.888080>
6. Chen, Q. (2025). Students' Perceptions of AI-Powered Feedback in English Writing: Benefits and Challenges in Higher Education. *International Journal of Changes in Education*. <https://doi.org/10.47852/bonviewijce52025580>
7. Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 29(11), 14151. <https://doi.org/10.1007/s10639-023-12402-3>
8. Fletcher, T., & Hayes-Birchler, A. (2023). Is remote measurement a better assessment of internet censorship than expert analysis? Analyzing tradeoffs for international donors and advocacy organizations of current data and methodologies. *Data & Policy*, 5. <https://doi.org/10.1017/dap.2023.5>

9. Gong, K. (2023). Challenges and opportunities for spoken English learning and instruction brought by automated speech scoring in large-scale speaking tests: a mixed-method investigation into the washback of SpeechRater in TOEFL iBT. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1). <https://doi.org/10.1186/s40862-023-00197-2>
10. Gupta, S., Unnam, A., Yadav, K. S., & Aggarwal, V. (2024). Towards Building a Language-Independent Speech Scoring Assessment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21), 23200. <https://doi.org/10.1609/aaai.v38i21.30366>
11. Handley, Z., & Wang, H. (2023). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency? *Language Assessment Quarterly*, 21(1), 3. <https://doi.org/10.1080/15434303.2023.2283839>
12. Ifenthaler, D. (2022). Automated Essay Scoring Systems. In *Handbook of Open, Distance and Digital Education* (p. 1). [https://doi.org/10.1007/978-981-19-0351-9\\_59-1](https://doi.org/10.1007/978-981-19-0351-9_59-1)
13. Javed, U., & Nabi, I. (2022). Heterogeneous fragility in Pakistan. In *Routledge eBooks* (p. 141). Informa. <https://doi.org/10.4324/9781003297697-5>
14. Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5, 100145. <https://doi.org/10.1016/j.cmpbup.2024.100145>
15. Kim, J., Chapman, M., Willner, L. S., Kemp, J. A., & Kim, A. A. (2026). Educator perspectives on automated writing scoring and feedback for young language learners: Applying a fairness and justice lens. *Assessing Writing*, 69, 101050. <https://doi.org/10.1016/j.asw.2026.101050>
16. Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12878-7>
17. Landa-Blanco, M. (2026). Artificial intelligence in education: applications and limitations for teachers in low- and middle-income countries. *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1681836>
18. Li, Y., Shan, Z., Raković, M., Guan, Q., Gašević, D., & Chen, G. (2025). When AI explains in natural language: Unveiling the impact of generative AI explanations on educators' grading and feedback practices. *Education and Information Technologies*, 30(17), 24931. <https://doi.org/10.1007/s10639-025-13741-z>

19. Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education Artificial Intelligence*, 6, 100234. <https://doi.org/10.1016/j.caeai.2024.100234>
20. Payande, I., & Charkameh, H. (2025). From .com to .gov: The internet's inevitable nationalist turn. *Internet Policy Review*, 14(3). <https://doi.org/10.14763/2025.3.2029>
21. Sajja, R., Sermet, Y., Fodale, B., & Demir, İ. (2026). Evaluating AI-powered learning assistants in engineering higher education with implications for student engagement, ethics, and policy. *Scientific Reports*, 16(1). <https://doi.org/10.1038/s41598-026-39237-5>
22. Saleem, T., Saleem, A., & Aslam, D. M. (2025). Integrating AI in Pakistani ESL classrooms: Teachers' practices, perspectives, and impact on student performance. *PLoS ONE*, 20(9). <https://doi.org/10.1371/journal.pone.0333352>
23. Sharma, P., Sharma, R. K., Singh, K., Maity, M., & Chakravarty, S. (2023). Dolphin: A Cellular Voice Based Internet Shutdown Resistance System. *Proceedings on Privacy Enhancing Technologies*, 2023(1), 589. <https://doi.org/10.56553/popets-2023-0034>
24. Uto, M., & Aramaki, K. (2024). Linking essay-writing tests using many-facet models and neural automated essay scoring. *Behavior Research Methods*, 56(8), 8450. <https://doi.org/10.3758/s13428-024-02485-2>
25. Yang, K., Raković, M., Li, Y., Guan, Q., Gašević, D., & Chen, G. (2024). Unveiling the Tapestry of Automated Essay Scoring: A Comprehensive Investigation of Accuracy, Fairness, and Generalizability. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20), 22466. <https://doi.org/10.1609/aaai.v38i20.30254>
26. Zellou, G., & Holliday, N. (2024). Linguistic analysis of human-computer interaction. *Frontiers in Computer Science*, 6. <https://doi.org/10.3389/fcomp.2024.1384252>
27. Zhang, S. (2021). Review of automated writing evaluation systems. *Journal of China Computer-Assisted Language Learning*, 1(1), 170. <https://doi.org/10.1515/jccall-2021-2007>
28. Zheng, A. (2024). Dissecting bias of ChatGPT in college major recommendations. *Information Technology and Management*. <https://doi.org/10.1007/s10799-024-00430-5>