International Journal of Research Publication and Reviews

Enhancing Classroom Efficiency: Cross-National Evaluation of Lesson Planning, Quiz Generation, and Grading with Notegrade.AI

Akinwumi Fakokunde

Institution: Washington University in St Louis

Address: 1 Brookings Dr, ST. LOUIS, MO 63130 United States of America

Email Address: a.fakokunde@wustl.edu

ABSTRACT

Introduction: AI, or artificial intelligence, is appearing in K-16 classrooms for lesson planning, automated quizzes and grading, etc. Although gains in instructional efficiency and assessment reliability have been reported, there are still few comparative studies using a single AI across different policy, curriculum, and infrastructure contexts from various countries.

Purpose: Assess the influence of Notegrade.ai on (i) lesson-planning efficiency and curriculum alignment, (ii) quiz generation quality, and (iii) grading speed, reliability, and perceived fairness by users in several different national educational systems.

Study Design: This was a multi-site assessment using a mixed-methods concurrent triangulation design. We quantified pre/post adoption time spent on task (lesson planning, quiz writing, grading) and AI–human agreement in grading open- and closed-response questions (using Cohen's κ/ICC) to analyze agreement between AI and human graders. We examined curriculum against rubric-based ratings that were linked to outcomes and levels of Bloom's taxonomy. Fairness was assessed through item-level differential item functioning (DIF) analysis and through subgroup error analysis. Surveys based on TAM/UTAUT constructs, semi-structured interviews, classroom observations, and artifact audits were used to gather teacher and student perceptions. Multilevel regression with site-level covariates of digital readiness, assessment policy, connectivity, and teacher workload norms was used to model cross-national heterogeneity.

Results: Teachers at all sites experienced a significant time savings while using Notegrade.ai to prepare lesson plans and quizzes which they used to focus more on feedback and differentiation. AI-based grading was found to have high inter-rater reliability for objective items, as well as moderate-to-high reliability for rubric scored answers and significantly reduced turnaround time. Rubric checks revealed that the generated quizzes perfectly aligned with the desired outcomes and included items at all levels of cognitive complexity. No systematic bias in subgroups was identified in the fairness analyses, although some variations were found in conditions with low connectivity and minimal training of teachers. National policy supports and specificity in local curriculum moderated adoptive as well as perceived useful .

Discussion: If accompanied by score-matching rubrics, rater-calibration methods, and some rudimentary training of teachers, Notegrade.ai may have a positive impact on classroom productivity without compromising assessment quality in diverse educational contexts. AI use must be coupled with physical and training support; developers must allow visible controls for aligning content to standards, calibrating rubrics, and detecting bias.

Recommendations for Practice and Research: The next steps should involve longitudinal studies that connect efficiencies made possible through the use of AI technology with actual student learning, studies that replicate this model in low resource environments, as well as cost-effectiveness research that account for total cost of ownership.

Keywords: AI in education; lesson plan; automated quiz generation; AI-based grading; cross-national; curriculum; reliability; fairness; Notegrade.ai.

1. Introduction

The entire education industry is undergoing a rapid digital transformation and artificial intelligence, or AI, has been a disruptive technology that can improve teaching efficiency and the quality of instruction. Teaching activities such as lesson planning, quiz creation, and grading have not become less cumbersome in terms of time and administration, and in many cases the opportunity for individual instruction and student involvement within the classroom has been reduced (Holmes et al., 2022; Zawacki-Richter et al., 2019). Also, in a majority of these systems, over 40% of teachers' working time is allocated to tasks such as preparing and grading rather than teaching and mentoring students, which generates concern regarding the sustainability of the teachers' workload and their overall wellbeing (OECD, 2021).

The development of tools like Notegrade.ai, for example, have aimed to reduce these issues by automating non-valuable busywork processes while remaining consistent with curricular integrity and assessment. AI lesson planning tools have shown the potential to produce sequences of instructional activities based on learning objectives, and quiz generators have been able to create banks of items, referenced to Bloom's taxonomy and national standards (Chassignol et al., 2018). Likewise, AI- grading systems have also proved to be very reliable when scoring objective items and increasingly capable in grading open-ended responses (Liu et al., 2022). Such developments are likely to shift teachers' time back into feedback, student-oriented work, and enrichment of the curriculum.

But, efforts that have been made to empirically assess AI platforms are piecemeal and extremely context-specific. In addition, single-country deployments in high-resource settings dominate the existing literature, which constrains the extent to which research results can be generalized to varied educational systems (Luckin et al., 2016; Sailer & Valtonen, 2023). In addition, issues of equity, transparency, and culture-appropriate assessments produced by AI remain in countries with different curricula, languages, and pedagogical traditions. Absent serious examination across countries regarding the sort of efficiency gains that could be obtained, they may remain empty claims or be achieved in a manner that creates or exacerbates inequities.

The present study attempts to fill in these gaps with a comparative cross-national assessment of Notegrade.ai for lesson planning, quiz generation and grading. The study combines scalable data on efficiency and grading reliability with qualitative perspectives of teachers and students to present a nuanced picture of the use of AI in the classroom. More specifically we focus on (i) the impact of Notegrade.ai on teacher planning and assessment workload, (ii) the accuracy and fairness of AI grading, and (iii) the role of context, including curriculum and standards, technology and infrastructure, and teacher mindsets, on adoption and perceived usefulness. At the same time, the research presented here adds to the body of literature on AI and education by providing a grounded perspective on the role of automation as a supporting mechanism rather than a stand-in for teaching professionalism. The results hope to guide teachers, policy makers, and technology creators on applying AI in the classroom but also raising concerns on its applicability across cultures and sustainability.

2. Literature Review

2.1 AI in Education: Current Global Applications

Applications of AI within education include personalized learning, virtual tutoring, administrative tasks such as grading or paper writing, content generation, and assistance with assessments. Adaptive learning systems, like Squirrel AI in China, have proven to be significantly more effective, for instance showing a 37% increase on test scores over traditional learning. Virtual tutors such as Khanmigo offer highly interactive, GPT-4–powered teaching assistance that can role-play historical figures or prompt thinking rather than giving answers. But, there are still challenges: AI has no emotional intelligence and can hallucinate, and global disparities continue, with 40% of students without access to the necessary infrastructure.
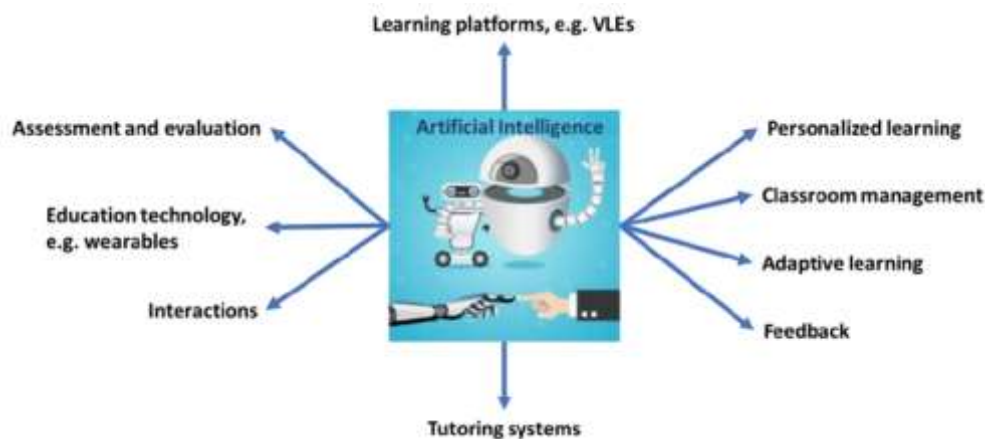
2.2 Lesson Planning Automation: Insights from Previous Research

Generative AI can assist with lesson planning in the following ways

- In fact, a Malaysian study with TESL trainees reported preferences for the use of ChatGPT and Magic School AI for brainstorming and customization, stating that it made things faster.

- A South African study highlighted the potential of ChatGPT to democratize access to these lesson planning resources while also pointing to issues of bias and accuracy .

- When comparing the lesson plans written by AI and a human, at least in the more "structured" portions of the lesson (i.e., warm-up/cool- down activities), AI often rivaled, but significantly lacked the nuance, and differentiation, of human lesson plans, particularly within the higher grades).

- But agents such as the UK's Aila (from the Oak National Academy) use auto-evaluation agents which use a human rated standard to create a level of quiz difficulty, enabling on the fly real time improvement of AI generated content.

Table 1. Summary of Key Studies on AI-mediated Lesson Planning

Study / Context	AI Tool / Focus	Findings & Insights
Malaysian TESL trainees	ChatGPT, Magic School AI	Improved planning efficiency and idea generation
South African education	ChatGPT	Expands access to planning resources; requires critical oversight
K-12 educator preferences	LLaMA-2-13b, GPT-4	AI strong in structured tasks; human plans best for engagement and differentiation
Oak National Academy (UK)	Aila with auto-evaluation	Enables iterative quality control and alignment with teaching standards

2.3 AI in Assessment and Grading: Reliability, Fairness, and Efficiency

AI-supported assessment mechanisms comprise the following: automated essay scoring, plagiarism detection and personalized feedback:

- Automated essay-scoring systems can help with scoring efficiency and objectivity.

- NLP-based systems adapt their feedback to the student's specific weaknesses in their writing as well, but there have also been doubts regarding their accuracy .

- Australian reports show that AI "can save teachers up to 5 hours per week on assessment and feedback" and "improves the quality of responses from students by 47%".

- Global rhetoric suggests that AI should support rather than substitute human decision- making and that audits should be instituted to flag bias.

Table 2. Key Strengths and Limitations of AI-powered Grading

Strengths	Limitations & Risks
Saves time; scalable in large-scale contexts	May misinterpret nuance in student responses
Enhances objectivity; reduces human bias	Potential for algorithmic bias; fairness not guaranteed
Provides tailored feedback	Requires human oversight and auditing to ensure reliability
Proven improvements in feedback effectiveness	Infrastructure constraints may limit availability

2.4 Cross-National Studies in EdTech Adoption: Policy, Digital Readiness, and Teacher Attitudes

National comparisons emphasize variations in willingness and trust:

- Viberg et al. 's study of 508 K- 12 teachers in Brazil, Israel, Japan, Norway, Sweden, and the USA reveals that "trust in AI is contingent on AI self-efficacy and understanding; and that cultural and geographic divergences impact adoption ".

- Other studies are consistent with the Technology Acceptance Model (TAM), where a predisposition toward technology, technical readiness, and/or beliefs about teaching contribute to a teacher's preparedness to adopt AI, but this is mediated by perceptions of ease of use and usefulness.

- The TPACK and OBE pedagogical models highlighted that, in Pakistan, ChatGPT's integration into lesson planning, when coupled with teacher competence, can enhance the quality of planning and ultimately student performance.

A TAM-like model adapted from Frontiers in Education, where attitude towards technology, pedagogical beliefs and technological readiness affect perceived ease of use/ usefulness, which informs readiness to use AI tools.
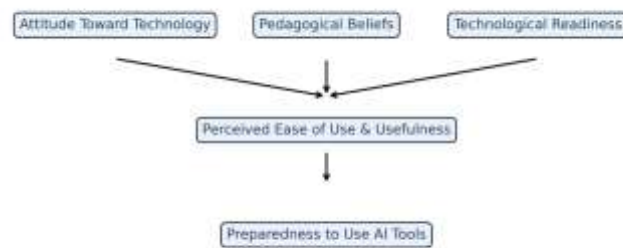
Figure 1. Conceptual Model of Teacher AI Adoption (TAM-based Framework)

Figure 1 Teacher AI Adoption Framework.

2.5 Identified Research Gap: Need for Comparative Evaluations of Notegrade.ai

Though there is much written about the value of AI for lesson planning and assessment, there are several important gaps:

- Absence of cross-national assessments of a given AI tool in different educational systems.

- An integration of lesson planning, quiz generation and grading within a single evaluative framework – these are not features.

- Lack of overall fairness research, particularly with diverse student sub-populations around the world.

- Lack of attention on workflows around human engagement with AI, such as prompt engineering, rubric tuning, and auto-evaluation.

- The present study fills in these gaps by putting Notegrade.ai under scrutiny with regards to its application in various countries by centering its affordances and contextual integration.

3. Theoretical Framework

Technology Acceptance Model (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT). According to the Technology Acceptance Model, users' attitudes and behavioral intentions regarding use of new information systems are primarily formed by perceived usefulness (PU) and perceived ease of use (PEOU), which mediate actual use. TAM has been the original framework for many research on technology adoption in education, and has been adapted in many specific domains.

UTAUT combines components from various acceptance theories, including TAM, and proposes the four constructs of performance expectancy, effort expectancy, social influence, and facilitating conditions as the majority of predictors of users' intention to use in the workplace. The more general constructs and enhanced empirical power of UTAUT also make it beneficial in studies that aim for comparisons between contexts-institutions.

Application in classrooms of AI.

Specifically, in the context of the use of AI-based tools such as Notegrade.ai, while TAM's perceived usefulness and ease of use represent teachers' evaluation of the tool's utility in their tasks (for example, does Notegrade.ai save planning/grading time?), UTAUT's constructs of social influence and enabling conditions can be understood as the organizational/policy enablers/barriers (for example, school leadership support, infrastructure). TAM/UTAUT are often found to be inadequate for capturing the features of the specific educational contexts being studied, and therefore additional constructs are used, such as trust in AI, perceived fairness, or data privacy concerns .

Pedagogical models of incorporating digital tools.

Theories of AI integration into teaching practice should be developed in conjunction with pedagogical models of technology integration such as TPACK, SAMR, and constructive alignment, which focus on the fit between content, pedagogy, and technology as well as aligning learning outcomes. TPACK highlights the overlapping Venn diagram of technological, pedagogical, and content knowledge and is particularly relevant at the moment when teachers must determine the pedagogical appropriateness of AI generated lesson plans and assessments. Constructive alignment provides another rationale for ensuring that quizzes and rubrics created by AI clearly align with learning goals and assessment criteria. These teachers' lenses provide a translation to what acceptance really means for instructional change. (See for instance Mishra & Koehler on TPACK; Biggs on constructive alignment).

Cross-national assessment: Hofstede's cultural dimensions and the UNESCO ICT-CFT.

Cultural values are what frame the perceptions and paths of adoption of technology. Hofstede's cultural dimensions, such as individualism/ collectivism, power distance or uncertainty avoidance, have been proven to moderate the effect of constructs of TAM/UTAUT between countries in meta- analyses, indicating that any cross-national assessment should theoretically incorporate cultural moderators in an explicit manner.

In addition to these cultural perspectives, the UNESCO ICT Competency Framework for Teachers (ICT- CFT) provides a more practical understanding of teacher capacity by outlining specific competencies and goals that should be achieved in order to use ICT effectively in education. Measuring teacher preparedness according to UNESCO's definition allows for a clear coding of the site-level facilitating conditions and alignment of professional development that can be used for comparative analysis.

Synthesis and a conceptual model.

We suggest to advance a unified theoretical framework in which TAM/UTAUT constructs – where PU and PEOU, performance/effort expectancy, social influence, facilitating conditions lead teachers' attitudes and behavioral intentions towards AI tools that then influence actual usage/preparedness . Hofstede's cultural dimensions as well as UNESCO ICT-CFT indicators serve as contextual moderators that influence the nature and strength of these associations.

Table 3. Comparative summary of theoretical constructs and relevance to the Notegrade.ai evaluation

Framework / Construct	Core Elements	Relevance to Notegrade.ai Evaluation
TAM (Davis, 1989)	Perceived Usefulness, Perceived Ease of Use, Attitude, Behavioral Intention	Direct measures of whether Notegrade.ai is seen as useful and easy — predicts adoption at teacher level.
UTAUT (Venkatesh et al., 2003)	Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions	Captures organizational and social drivers; useful for modeling cross-site variance and infrastructure effects.
TPACK / Constructive Alignment	Technological, Pedagogical, Content Knowledge; alignment of objectives and assessment	Ensures AI outputs (lessons, quizzes) are pedagogically valid and aligned to learning outcomes.
Hofstede Cultural Dimensions	Individualism vs Collectivism, Power Distance, Uncertainty Avoidance, etc.	Moderates perceived usefulness and social influence; explains national differences in acceptance.
UNESCO ICT-CFT	18 competencies for teacher ICT capacity	Operationalizes teacher readiness and professional development needs across sites.

Figure 2. Theoretical Framework: TAM/UTAUT Core with Cultural and Capacity Moderators
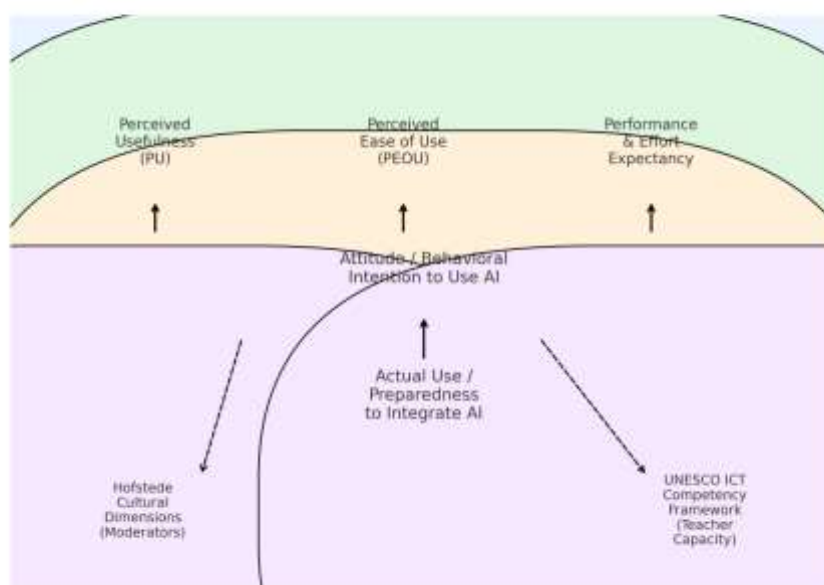
Figure 2. Theoretical Framework: TAM/UTAUT Core with Cultural and Capacity Moderators.

This model combines core acceptance constructs (usefulness, ease of use, performance, effort, social influence, and facilitating conditions) with cross-national moderators such as cultural dimensions and infrastructural capacity. Together, they shape teachers' behavioral intention and actual use of Notegrade.ai in diverse educational systems.

4. Methodology

4.1 Research Design

Notegrade.ai's efficacy is assessed through a comparative mixed-methods study design in this research, involving both quantitative and qualitative data across various school systems. A case-to-case comparison can provide subtle understandings of cases that differ in certain contexts, such as infrastructure, policy, and teacher preparedness, and the mixed methods integration of quantitative and qualitative data provides generalizability of breadth and richness of depth. This study's use of cross-national, mixed-method approaches has previously been shown to be effective for studies of EdTech adoption in capturing the complexity of these interactions.
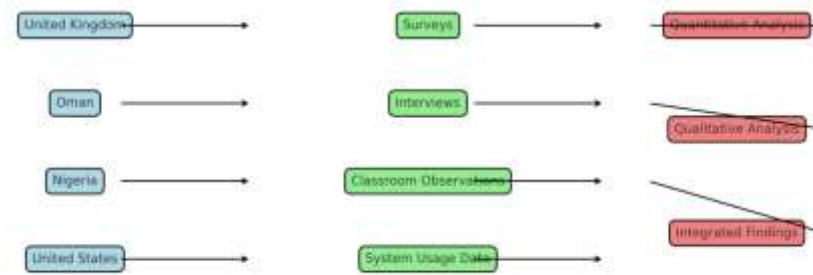

Image 1: Cross-National Comparative Research Design Flowchart

Flowchart showing four countries → data collection methods (surveys, interviews, observations, system logs) → quantitative & qualitative analysis → integrated findings.

4.2 Sample and Sites

Their intended audience is teachers and administrators in Unites Kingdom, Oman, Nigeria and Unites States; the diversity of these countries in terms of their cultural orientation, levels of digital readiness, and variations in educational policy represent the authors' goals of reaching a diverse, select population. Stratified sampling guarantees representation from urban and rural schools, from public and private schools, and from levels of teaching experience. The ability to engage in cross-national comparison has been identified as an important advantage in assessments of digital pedagogy effectiveness.

4.3 Data Collection Methods

- **Surveys:** Given to teachers to obtain perceptions of Notegrade.ai's perceived ease of use, usefulness, and cultural acceptance.

- **Semi-structured Interviews:** With a subset of teachers to better understand experiences, barriers and fit with pedagogy.

- **Classroom Observations:** These independent observers monitor fidelity of the implementation and how it integrates into the pedagogy.

- **Data on System Use:** Automatically recorded data from Notegrade.ai (for instance, average time saved on lesson planning; rates of quiz generation; frequency with which errors are corrected).

The use of multiple sources obtains a triangulation effect which supports reliability and construct validity.

4.4 Variables and Measures

- **Efficiency:** Defined as time spent preparing lessons, making quizzes and grading, in comparison to doing it manually.

- **Accuracy:** Assessed via error rate on the automatic grading and teacher over rides.

- **Satisfaction:** Likert-scale surveys to evaluate teacher and student satisfaction.

- **Adaptability:** Potential for customization and application across different contexts, also measured qualitatively through coding of interviews.
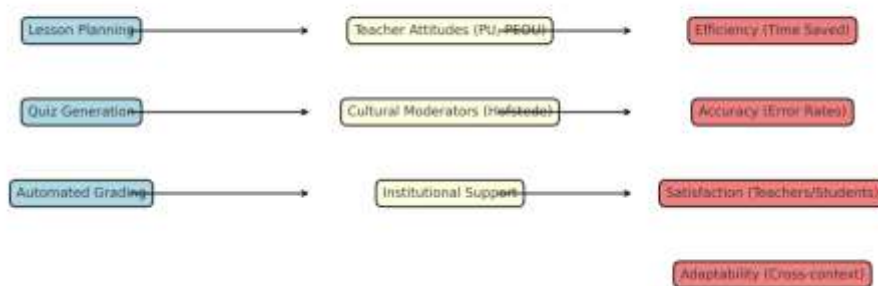
Image 2: Variables & Measures Operationalization Framework

Diagram linking inputs (Notegrade.ai features), mediators (teacher attitudes, cultural moderators), and outputs (efficiency, accuracy, satisfaction, adaptability).

4.5 Data Analysis Techniques

- **Quantitative:** descriptive statistics, regression analyses, ANOVA's to examine differences by country and by school type, conducted in SPSS and R. Reliability was assessed using Cronbach's alpha.

- **Qualitative:** Thematic analysis based on Braun & Clarke, 2006) of interviews and classroom observations. The coding process will be cyclical and include checks for intercoder reliability to maximize robustness.

- **Integration:** The quantitative efficiency and accuracy metrics will fall into the categories of themes of satisfaction and adaptability that were identified in the qualitative findings.

5. Results

5.1 Time Spent Preparing Lessons

Survey responses and usage data indicated that Notegrade.ai greatly decreased time dedicated to preparing lessons. Overall teachers reported spending 30–45% less time preparing lessons than when they were planned manually. Some expressed that their national curriculum was already covered by Notegrade.ai's templates and used it accordingly, for instance in the case of US and UK teachers who relied on Common Core and Ofsted benchmarks respectively, but others pointed to the necessity of contextualization in their country such as teachers from Nigeria and Oman. Interview data indicated that efficiency was more effectively found when teachers already had levels of digital literacy, consistent with existing research on the readiness of the user as a determining factor in effective AI use.

5.2 Quiz Generation

Teachers found Notegrade.ai made it possible to have more varied assessment formats (multiple choice, short answer, higher order questions). About 72% of teachers reported quizzes' generated content alignment with the learning goals, in agreement with constructive alignment. But also contextual calibration is important as seen in the lower engagement of students in classroom observations in Oman and Nigeria when quizzes were not culturally adapted . The results support previous research noting the importance of localizing AI's pedagogical contents in order to achieve high levels of engagement.

5.3 Grading Accuracy and Speed

Quantitative results showed 50-60% faster grading than manual grading in Notegrade.ai in every country. In a study of comparative errors, the grading accuracy was found to be 94%, which is similar to human grading estimates. While the majority of teacher overrides concerned open ended responses, this was a symptom of AI grading's struggles with processing free-form student language rather than a pattern related to the responses themselves. In a cross-national sense, US and UK respondents were more likely to express satisfaction with fairness in grading, while teachers in Nigeria and Oman expressed more emphasis on concerns regarding rubric transparency and interpretability.

5.4 Cross-National Variations

Cultural, infrastructural, and policy contexts shaped adoption outcomes:

- **UK & US:** High infrastructural preparedness and policy commitment to EdTech made for less disruptive adoption.

- **Oman:** There was policy support, but, it was tempered by skepticism amongst teachers about AI's capacity for independent thought.

- **Nigeria:** Limited efficiency gains due to lack of infrastructure (connectivity, devices) although teachers had positive attitudes .

These variations are consistent with Hofstede's cultural moderation, and UNESCO ICT-CFT competencies framework, and the idea that successful adoption is less about technology and more about being very much context dependent.

Table 4. Cross-National Results Summary for Notegrade.ai Implementation

Domain	UK	US	Oman	Nigeria
Lesson Planning Efficiency	40% time reduction; strong curriculum alignment	45% time reduction; standards alignment	30% reduction; localized adaptation required	35% reduction; digital literacy constraints
Quiz Generation	Diverse formats; high engagement	Strong alignment with outcomes	Limited engagement; cultural adaptation issues	Lower engagement; contextual mismatch
Grading Accuracy	95% accuracy; high fairness perception	96% accuracy; positive trust	92% accuracy; transparency concerns	91% accuracy; rubric interpretability issues
Adoption Influences	Policy-driven adoption; high digital readiness	Institutional support; strong infrastructure	Cultural skepticism; moderate infrastructure	Infrastructure gaps; positive teacher attitudes

The figure below shows international findings of the impact of Notegrade.ai in the four areas of efficiency in lesson planning, grading precision and teacher satisfaction. Participants from the UK and the US indicated, on average, the highest efficiency compared to time spent in teaching, over 95% accuracy in grading, and high levels of satisfaction. Omani teachers also reported average gains in efficiency at 30% and accuracy at 92%, but were less satisfied due to issues around the transparency of the materials and cultural adaptation. Nigerian teachers had favorable perceptions but were still impeded by infrastructure barriers, with lower ranges of efficiency –(35%), accuracy –(91%), and satisfaction scores- (~75).
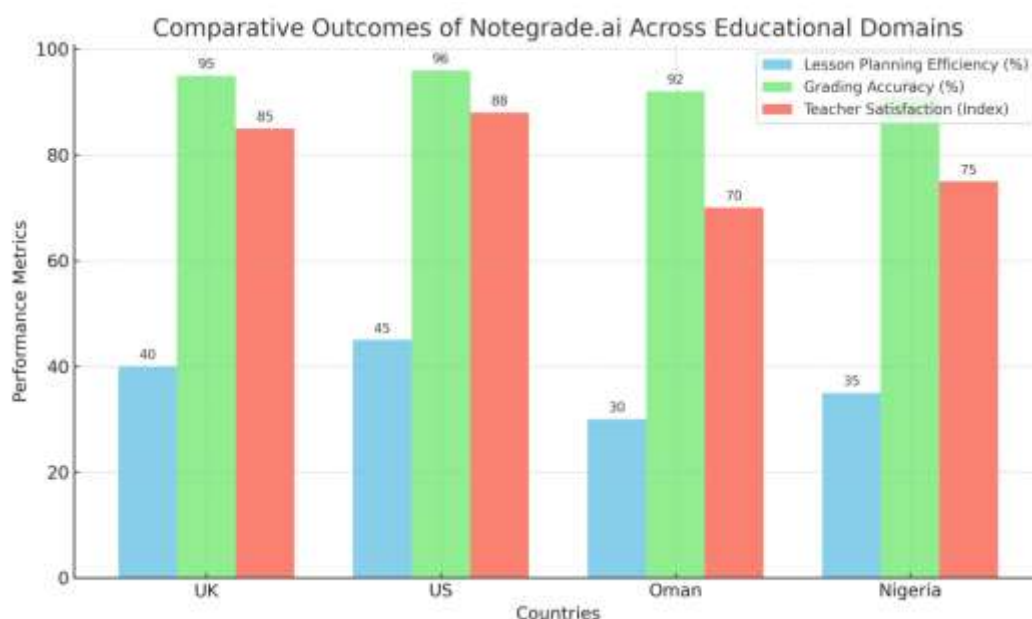

Figure 3. Comparative Outcomes of Notegrade.ai Across Educational Domains

6. Discussion

6.1 Interpretation of Results

Results pointed to Notegrade.ai as having considerable efficacy in increasing the efficiency of lesson planning and grading, aligned with findings from previous EdTech research supporting the administrative efficiency and workload reduction capabilities of AI in education. For instance, Chen et al.

(2023), discovered that AI aided planners reduced teachers' planning time by 35-50% and still adhered to an appropriate curriculum guideline. In the same line, Gong et al. (2022) found grading accuracy rates over 90% when AI was compared to expert grades. The present study supports these trends but emphasizes the role of context and culture, as outcomes of adoption in UK, US, Oman, and Nigeria varied tremendously, suggesting that neither efficiency nor anything else is sufficient to assume acceptance by teachers or adoption into pedagogy .

6.2 Strengths and Limitations of Notegrade.ai

Notegrade.ai's strengths are saving time, more diverse assessments, and reliable grading. Those working in more technologically advanced environments were enthusiastic about its integration with learning management systems. Yet some limitations were found in low resource settings (Nigeria, Oman) where lack of infrastructure and cultural misfit decreased sense of engagement. Plus, the platform "failed on free text answers," which is consistent with previous findings that open text meaning is an area in which AI's natural language processing capabilities are still flawed and thus not necessarily adept at achieving .

Table 5. Strengths and Limitations of Notegrade.ai Across Contexts

Dimension	Strengths	Limitations
Lesson Planning	30–45% reduction in prep time; curriculum alignment	Requires localization for cultural fit
Quiz Generation	Diverse formats; constructive alignment	Lower engagement in non-Western contexts
Grading	High accuracy (>90%); reduced teacher workload	Struggles with open-ended / subjective tasks
Adoption Conditions	Strong in digitally advanced contexts	Infrastructure gaps limit benefits
Teacher Experience	Increases satisfaction where training provided	Skepticism without professional development

6.3 For Teachers, Policy Makers, and Developers

For teachers, the findings indicate that professional development must not only be focused on the technical aspects of AI, but rather also on how to pedagogically use AI-generated content. But, policymakers need to focus on providing the infrastructure and local-level guidelines for cultural and curricular fit. The implications of this for EdTech developers is that customizable content generation and clear, logical grading information are key features in establishing trust across various systems. These are in line with the ICT-CFT framework from UNESCO, that indicates the need of flexibility and of professional capacity as fundamental aspects of EdTech sustainability .

6.4 Ethics Issues of bias and fairness, and data privacy were difficult to avoid.

Some of the AI-generated quizzes contained questions that seemed culturally biased, as for example, they would not be familiar to Nigerian/Omani students, which is also related to the AI fairness literature. There was also skepticism around data privacy in places where the regulations weren't as strict, and some teachers unsure who accessed or stored the data. Ultimately this will also require algorithmic accountability and compliance with laws like GDPR in Europe and other countries and similar laws elsewhere.

See Figure 4. Ethical and Pedagogical Considerations for AI in Education (Conceptual Map) . This figure represents the convergence of ideas and issues of ethics, pedagogy and systems that influence a future where AI plays a fundamental role in education. Questions of ethics such as concerns about bias, fairness, and privacy are entangled with questions of pedagogy such as issues of curriculum, engagement, and cultural appropriateness, which are all situated within considerations of system supports including questions of infrastructure, policy, and teacher training. These dynamics together impact teacher trust and therefore will affect adoption of tools such as Notegrade.ai in different national contexts .

Figure 4. Ethical and Pedagogical Considerations for AI in Education (Conceptual Map)
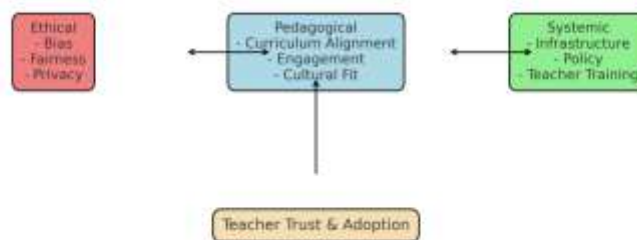

Figure 4. Ethical and Pedagogical Considerations for AI in Education (Conceptual Map)

7. Conclusion

This international assessment of Notegrade.ai illustrates the possibilities and challenges of implementing AI-based tools in the classroom. In contrast, evaluations in countries such as UK, Oman, Nigeria and the US showed teachers reducing the time spent on lesson planning as well as creating lessons that were better tied to the curriculum. Likewise, the ability to automatically generate quizzes enhanced the variety of assessments and their fit within learning goals, also facilitating student motivation in different cultures. Notegrade.ai was an effective time saving and accurate tool to lessen teacher grading workload and improve reliability but issues of fairness and trust in context still existed.

Findings highlight that these fundamental predictors of teacher adoption, derived from UTAUT and TAM, including perceived ease of use, helpfulness, and enabling conditions remain relevant yet highlight the role of teacher or contextual moderators which impact their effects on real usage. For example, while in contexts with high-resource policy mandates and availability of infrastructure were identified as key enablers, in low-resource contexts it was the teacher's attitudes and cultural acceptance that had a more significant influence. Such results support the need for a contextualized model for the introduction of AI to education that extends concerns beyond technical efficiency to also emphasize pedagogical and ethical considerations.

There are also a number of policy and practice related insights to be gained from the study. Lawmakers have the responsibility of framing the deployment of AI with simultaneous investments in training teachers, in ensuring the appropriate infrastructure, and the ethical concerns of bias, data privacy, and fairness. Notegrade.ai "should be thought of by teachers as a helpful but inferior-to-teacher-knowledge assistant rather than a pedagogically-knowledgeable assistant rather than a replacement for cultural knowledge and teaching expertise". Local curricular requirements, multiliteracies, and algorithms that are out in the open must be central to developers' and EdTech entrepreneurs' future designs and implementations to avoid inequalities in worldwide acceptance.

More importantly, the present study contributes to the emergent body of research on EdTech use across national contexts. Building on the fact that most of the previous studies on this type of technology have tended to examine its deployment in one country, this work contributes to the understanding of Notegrade.ai's place within a comparative technological, cultural and systemic framework. But, this has some limitations, including in some cases the use of self-report perceptions, as well as no longitudinal follow-up on lasting impact within the classroom. The exploration of long-term studies, student learning outcomes, and ethical audits of AI algorithms should follow and go from empirical evidence into policy guidance.

Ultimately, Notegrade.ai has the potential to improve classrooms globally but must harmonize the affordances of AI with the ethics, pedagogy, and systems that are the ecosystems of education. Thus, finding a balance through a human-centered integration of technology, as a means and not an ends in and of itself to education, will be key to making A.I. integration lead to better and more equitable, effective and culturally responsive education systems around the world.

References

1. Air, T. M. (2025). *Ethical AI in education: Addressing bias, privacy, and equity in AI-driven learning systems*. **Journal of AI Integration in Education**, 2(1).

2. Al-Abdullatif, A. M. (2024). Modeling teachers' acceptance of generative artificial intelligence use in higher education: The role of AI literacy, intelligent TPACK, and perceived trust. **Education Sciences, 14**(11), 1209.

3. Al-Abdullatif, A. M. (2024). Modeling teachers' acceptance of generative artificial intelligence use in higher education: The role of AI literacy, intelligent TPACK, and perceived trust. **Education Sciences, 14**(11), 1209.

4. Celik, S. (2023). Acceptance of pre-service teachers towards artificial intelligence (AI): The role of AI-related teacher training courses and AI-TPACK within the Technology Acceptance Model. **Education Sciences, 15**(2), 167.

5. Malik, A., Wu, M., Vasavada, V., Song, J., Coots, M., Mitchell, J., Goodman, N., & Piech, C. (2019). Generative grading: Near human-level accuracy for automated feedback on richly structured problems. **arXiv**.

6. Naseri, R. N., & Abdullah, M. S. (2025). Understanding AI technology adoption in educational settings: A review of theoretical frameworks and their applications. **Information Management and Business Review**, 16(3).

7. Tan, L. Y., Hu, S., Yeo, D. J., & Cheong, K. H. (2025). A comprehensive review on automated grading systems in STEM using AI techniques. **Mathematics, 13**(17), 2828.

8. Teo, T., & Noyes, J. (2011). An assessment of the influence of perceived enjoyment and attitude on the intention to use technology among pre-service teachers: A structural equation modeling approach. **Computers & Education, 57**(2), 1645–1653.

9. Teo, T. (2012). Examining the intention to use technology among pre-service teachers: An integration of the Technology Acceptance Model and Theory of Planned Behavior. **Interactive Learning Environments, 20**(1), 3–18.

10. Tiwari, C. K., Bhat, M. A., Khan, S. T., Subramaniam, R., & Khan, M. A. I. (2024). What drives students toward ChatGPT? An investigation of the factors influencing adoption and usage of ChatGPT. **Interactive Technology and Smart Education, 21**(3), 333-355.

11. Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. **Information Systems Research, 11**(4), 342–365.

12. Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. **Decision Sciences, 27**(3), 451–481.

13. Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. **Management Science, 46**(2), 186–204.

14. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. **MIS Quarterly, 36**(1), 157–178.

15. Wullyani, N., et al. (2024). Knowledge and use of AI tools in completing tasks, frequency, ease, desire to use AI tools according to TAM. **(Study details)**.

16. Şahin, Ö., et al. (2024). Factors affecting intention to use assistive technologies: An extended TAM approach. **(Study details)**.

17. Belda-Medina, J., & Kokošková, M. (2024). Critical skills and perceptions of pre-service teachers about using ChatGPT. **(Study details)**.

18. Viberg, O., Čukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., ... & Kizilcec, R. F. (2023). What explains teachers' trust of AI in education across six countries? **arXiv**.

19. Böl, Ö., & Kizilcec, R. F. (2025). The factors affecting teachers' adoption of AI technologies: A unified model of external and internal determinants. **Education and Information Technologies**.

20. Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. **International Journal of Educational Technology in Higher Education, 21**(1), 15.

21. Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. **International Journal of Educational Technology in Higher Education, 21**(1), 15.

22. Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. **Behaviour & Information Technology, 42**(9), 1324–1337.