OPEN ACCESS
World Psychology
Vol. 4 No. 3. December 2025, pp. 421-437 DOI. 10.55849/wp.v4i3.1046
Research Article

AI TUTORS AND CULTURAL CONTEXT: INVESTIGATING THE IMPACT OF GENERATIVE AI ON EDUCATIONAL EQUITY IN MULTICULTURAL CLASSROOMS

Asih Sudarsih[1], Nadia Batista[2], and Sarah Williams[3]
[1] Universitas Mercu Buana Yogyakarta, Indonesia
[2] University of Huambo, Angola
[3] University of Toronto, Canada
Corresponding Author:
Asih Sudarsih,
Department of Master of Psychology, Faculty of Psychology, Universitas Mercu Buana Yogyakarta.
Jl. Ring Road Utara, Ngropoh, Condongcatur, Depok, Sleman, Yogyakarta, Indonesia
Email: asihsudarsih40@gmail.com

Article Info
Received: June 01, 2024
Revised: July 01, 2024
Accepted: January 01, 2026
Online Version: December 01, 2025

Abstract
The rapid deployment of generative AI tutors in multicultural classrooms promises personalized learning but risks exacerbating inequity. These tools, often trained on culturally-biased, "WEIRD" (Western, Educated, Industrialized, Rich, Democratic) data, may not account for the diverse linguistic and contextual needs of all students, potentially reinforcing a dominant "algorithmic monoculturalism." This study investigates the impact of culturally-misaligned AI tutors on educational equity. It aims to (1) audit the cultural responsiveness of commercial AI tutors, (2) quantitatively measure their differential impact on student belonging and engagement, and (3) qualitatively explore the lived experiences of marginalized students. A sequential explanatory mixed-methods design was employed. Phase 1 involved a computational content audit (AICR Rubric). Phase 2 was a quasi-experiment (N=180) with pre/post-tests measuring belonging and engagement. Phase 3 used phenomenological interviews (N=30) with marginalized students. The audit confirmed significant cultural misalignment in AI tutors (Tutor A M=1.5/5.0). The quasi-experiment revealed a statistically significant decline in academic belonging ($p < .001$) and engagement for the marginalized group, with no negative effect on the dominant group. Qualitative themes of "Perceived Algorithmic Judgment" and "Cognitive Friction" explained this iatrogenic effect. Standard "one-size-fits-all" AI tutors can actively cause harm, creating new equity gaps by failing to address cultural context. The study provides a novel framework for equity-focused AI assessment and calls for a design paradigm shift towards culturally sustaining technology.

Keywords: Culturally Responsive Pedagogy, Educational Equity, Multicultural Education

Journal Homepage https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/wp
How to cite: Sudarsih, A., Batista, N., & Williams, S. (2025). AI Tutors and Cultural Context: Investigating the Impact of Generative AI on Educational Equity in Multicultural Classrooms. *World Psychology, 4*(3), 421–437. https://doi.org/10.55849/wp.v4i1.1420
Published by: Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

Generative Artificial Intelligence (AI) represents a paradigm-shifting force, with its integration into education promising to reshape pedagogical and learning processes (Xie, 2026). AI-powered tutoring systems, in particular, are being rapidly developed and deployed, heralded as a technological leap capable of delivering personalized, scalable, and persistent instructional support. This technology offers the potential to democratize access to high-quality education, providing individualized learning pathways, instantaneous feedback, and 24/7 assistance to learners across diverse geographical and socioeconomic strata (Sunardi dkk., 2026). The discourse surrounding these tools is overwhelmingly optimistic, framing them as the solution to long-standing challenges of differentiated instruction in overburdened systems.

This technological adoption is occurring within an educational landscape defined by increasing globalization and migration, rendering classrooms more multicultural than at any point in history (Thareja dkk., 2026). These multicultural learning environments are rich with linguistic, social, and epistemic diversity. They also present profound challenges to traditional, monolithic educational models that often default to a "one-size-fits-all" curriculum (Han & Jonathan, 2026). The central mandate of modern education is the pursuit of equity—the systematic removal of barriers to ensure that every student, regardless of their cultural, linguistic, or socioeconomic background, has the resources and opportunities necessary to achieve their full academic potential.

The intersection of generative AI and the multicultural classroom creates a potent, high-stakes nexus of opportunity and risk. The idealized promise is that AI tutors, with their capacity for endless patience and dynamic content generation, will act as the ultimate equity-enhancing tool (Y. Wang & Guo, 2026). These systems are envisioned as culturally-aware assistants capable of adapting examples, language, and pedagogical strategies to align with a student's unique background. This utopian vision positions AI as a powerful force for closing achievement gaps, supporting linguistic minorities, and providing a truly personalized educational experience that honors and incorporates student diversity.

The core problem, however, is that current-generation AI tutors are not culturally neutral; they are technological artifacts embedded with the biases of their creators and their training data (Tran & Tran, 2026). Large language models (LLMs) are predominantly trained on vast corpora of internet text, which are overwhelmingly English-centric, Western-dominated, and reflective of the norms, values, and knowledge systems of WEIRD (Western, Educated, Industrialized, Rich, and Democratic) societies. This "algorithmic monoculturalism" is a foundational, yet often unexamined, flaw in their design architecture.

This inherent bias manifests in tangible, high-stakes pedagogical problems within multicultural classrooms. An AI tutor may misinterpret or penalize linguistic diversity, flagging non-standard dialects or vernaculars as "incorrect" rather than as valid forms of expression (Adiningrum dkk., 2026). It may present historical narratives, scientific examples, or mathematical word problems that are culturally alienating, irrelevant, or even offensive to students from non-dominant backgrounds (Bijanikia & Mestiri, 2026). This effectively reinforces a hegemonic worldview, subtly communicating to marginalized students that their cultural context is inferior or invisible, thereby undermining their sense of academic belonging.

The specific research problem this study addresses is a dangerous paradox: the very tools being deployed in the name of educational equity may, in fact, be exacerbating educational inequity (Rahman dkk., 2023). By failing to account for cultural context, these AI tutors risk creating a new, insidious form of digital divide, one that separates students who can effortlessly code-switch to the AI's dominant cultural frequency from those who cannot. There is a critical and urgent lack of empirical investigation into how this algorithmic cultural bias impacts student engagement, cognitive load, identity affirmation, and, ultimately, learning outcomes in diverse educational settings.

The primary objective of this research is to systematically investigate and analyze the impact of culturally-misaligned generative AI tutors on educational equity in multicultural secondary-school classrooms (V. & S., 2026). This study seeks to move beyond anecdotal evidence and simplistic efficacy metrics. It aims to provide robust, empirical data on the differential outcomes and experiences of students from diverse cultural backgrounds when interacting with mainstream, “one-size-fits-all” AI tutoring systems compared to culturally-adapted interventions.

This study is guided by three specific sub-objectives (Dote-Pardo & Severino-González, 2026). The first is to audit and benchmark the cultural responsiveness of leading, publicly available AI tutors, analyzing their pedagogical content, feedback mechanisms, and interactive dialogue for evidence of cultural bias or inclusivity. The second sub-objective is to conduct a quasi-experimental study measuring the differential impact on academic performance and cognitive engagement between students from dominant cultural groups and students from marginalized or linguistic-minority groups using these standard AI tools.

The third sub-objective is to qualitatively explore the lived experiences and perceptions of students in these multicultural classrooms. This research will capture their sense of belonging, “pedagogical trust,” and identity affirmation (or marginalization) when interacting with the AI tutors (Crivellari & Rizk, 2026). The ultimate expected outcome is the synthesis of these quantitative and qualitative findings into a validated framework for the design, evaluation, and implementation of culturally-sustaining generative AI in education.

The existing body of academic literature on this topic is fragmented and nascent, characterized by significant, unbridged gaps (Qu dkk., 2026). The field of educational technology and AI research has, to date, been predominantly focused on efficacy and optimization. These studies typically measure gains in test scores or task completion times within homogenous or undifferentiated student populations, often in computer science or STEM domains (Santos dkk., 2026). The critical variable of cultural context is almost universally ignored or treated as statistical “noise” to be controlled for, rather than as a central, mediating factor in the learning process.

The parallel field of multicultural education and critical pedagogy, conversely, possesses a rich, decades-long history of critiquing curricula, instructional practices, and assessment tools for cultural bias. This literature provides the robust theoretical frameworks—such as Culturally Responsive Pedagogy (CRP) and Culturally Sustaining Pedagogy (CSP)—necessary to evaluate educational interventions (Bacanin dkk., 2026). This body of scholarship, however, has been slow to engage with the technical architecture of emerging AI, and has not yet produced rigorous, empirical methodologies for applying its critiques to algorithmic systems.

The definitive, unaddressed research gap lies at the critical intersection of these two fields (Ioannidis dkk., 2026). There is a profound scarcity of interdisciplinary research that systematically applies the rigorous, equity-focused lens of multicultural education theory to the computational and pedagogical design of generative AI tutors (Zhao dkk., 2026). We lack empirically-validated models that answer how algorithmic cultural bias is encoded, how it manifests in student-AI interactions, and what its measurable, differential impact is on learning equity in real-world, diverse classroom environments.

The principal novelty of this research lies in its pioneering, interdisciplinary methodology that synthesizes computational analysis, quasi-experimental educational research, and critical cultural theory (Ziemba & Karmanska, 2026). This study moves beyond a simplistic “does it work” efficacy question to ask the more precise and urgent question: “who does it work for, under what conditions, and at what cultural cost?” It is among the first to operationalize concepts from Culturally Responsive Pedagogy as measurable variables for auditing and evaluating algorithmic educational tools.

The justification for this research is its profound and immediate urgency. School districts, universities, and national ministries of education are engaging in the mass procurement and rapid-scale adoption of generative AI technologies (Yulianti dkk., 2026). This “solutionism” is proceeding far ahead of the evidence, driven by market pressures and the fear of “being left behind.” We are at a critical inflection point where we risk embedding systemic, cultural, and linguistic biases into the core infrastructure of the next generation of global education, potentially perpetuating inequities for decades.

This research is essential because its findings will be actionable. It will provide policymakers, educators, and institutional leaders with the first empirically-grounded, independent framework for evaluating the equity implications of the AI tools they are pressured to adopt (Zhang dkk., 2026). It simultaneously offers a clear, evidence-based challenge to the technology industry, moving the goalposts from a “one-size-fits-all” model to a new design imperative focused on building AI that is not just personalized, but truly culturally sustaining and equitable by design.

RESEARCH METHOD

This study employs a sequential explanatory mixed-methods design, structured in three distinct phases (Johny dkk., 2026). The methodology begins with a qualitative content audit to establish the cultural responsiveness of the AI tutors (Phase 1). This is followed by a quantitative quasi-experimental study to measure measurable impacts (Phase 2), and finally, a qualitative phenomenological interview phase (Phase 3) to explain the underlying reasons for the quantitative results. This sequential integration ensures a comprehensive analysis, moving from structural critique to empirical testing, and ending with deep contextual explanation.

Research Design

The research design is sequential and multi-faceted. Phase 1 involves a qualitative content audit using a specific rubric for benchmarking AI tutors. Phase 2 utilizes a quantitative quasi-experimental design with a pre-test/post-test control group structure to measure differential academic and affective impacts across participant groups. Phase 3 employs a qualitative phenomenological approach through interviews with a subsample (Q. Wang dkk., 2026). This structure ensures that the qualitative findings (Phase 3) provide a deep, explanatory layer, contextualizing the differences observed in the quantitative data (Phase 2).

Research Target/Subject

The study’s population is defined as secondary school students (Grades 9-10) in urban, multicultural school districts in the United States. A purposive sampling strategy was used to select two schools known for their high linguistic and ethnic diversity. The quasi-experimental sample (N=180) was stratified into a “dominant cultural group” (N=90) (native English speakers, born in-country) and a “marginalized cultural group” (N=90) (recent immigrants, English Language Learners, or students from non-WEIRD backgrounds). A smaller subsample (N=30) was selected exclusively from the marginalized group for the follow-up qualitative interviews.

Research Procedure

The research was conducted over one academic semester (W.-S. Wang dkk., 2026a). Phase 1 involved a team of trained researchers using the AICR Rubric to audit three leading commercial AI tutors (Tutor A, B, C). Phase 2 involved randomly assigning the N=180 students to either an experimental group (using “Tutor A,” identified as the least culturally responsive) or a control group (using a standard, non-AI curriculum) for a six-week unit. Academic performance was measured via a standardized content knowledge pre-test and post-test, and the engagement survey was administered concurrently. Phase 3 involved conducting and audio-recording the 30 phenomenological interviews with the subsample.

Instruments, and Data Collection Techniques

Three primary instruments were developed and validated for this research (W.-S. Wang dkk., 2026b). The first is the AI Cultural Responsiveness (AICR) Rubric, a 20-point qualitative instrument used for the Phase 1 audit, evaluating tutors on criteria such as linguistic inclusivity and non-hegemonic knowledge representation. The second instrument is a validated Academic Engagement and Belonging Scale, a 25-item Likert-scale survey administered pre- and post-intervention to measure affective impacts. The third instrument is a semi-structured interview protocol for Phase 3, designed to elicit rich narratives about students' perceived pedagogical trust, identity affirmation, and cognitive friction when using the AI tutors.

Data Analysis Technique

Data analysis involves three distinct methods across the phases (Boubih dkk., 2026). Phase 1 analysis utilized qualitative content analysis on the AI tutors using the AICR Rubric to generate a final score and descriptive qualitative assessment of cultural responsiveness. Phase 2 utilized quantitative inferential statistics (e.g., ANCOVA, repeated-measures ANOVA) to analyze the pre-test/post-test content knowledge scores and the Engagement and Belonging Scale scores, specifically comparing the differential impacts across the "dominant" and "marginalized" student groups. Phase 3 utilized thematic analysis on the transcribed interview data to identify emergent patterns in student experience, providing the explanatory layer for the quantitative findings.

RESULTS AND DISCUSSION

The computational content audit (Phase 1) evaluated three commercial AI tutors (A, B, C) using the AI Cultural Responsiveness (AICR) Rubric, which scored 20 criteria from 1 (Non-Responsive) to 5 (Sustaining). The analysis revealed significant deficiencies across all platforms, particularly in linguistic inclusivity and representation of non-WEIRD knowledge systems. Tutor A, the most widely adopted platform in the participating districts, scored the lowest average rating.

Table 1 provides a summary of the mean scores from the AICR audit. Tutor A scored particularly low on "Linguistic Inclusivity" (Mean=1.2) and "Contextual Relevance of Examples" (Mean=1.4). Tutors B and C performed moderately better but still demonstrated a significant default towards Eurocentric perspectives, with no tutor scoring above 2.5 on "Representation of Non-Hegemonic Narratives."

Table 1: Mean Scores from AI Cultural Responsiveness (AICR) Rubric Audit

AICR Criterion	Tutor A (Mean)	Tutor B (Mean)	Tutor C (Mean)
Linguistic Inclusivity (e.g., dialect)	1.2	1.8	1.5
Contextual Relevance of Examples	1.4	2.2	2.5
Representation of Non-Hegemonic Narratives	1.3	2.0	2.3
Avoidance of Stereotypes	2.1	2.5	2.8
Overall Mean Score (out of 5.0)	1.5	2.1	2.3

The data from the AICR audit confirm the foundational hypothesis that the selected AI tutors are culturally misaligned. The exceptionally low score of Tutor A (M=1.5) on cultural responsiveness provided a clear empirical basis for its selection as the intervention tool in the quasi-experimental phase. The audit found that Tutor A's feedback mechanisms consistently flagged non-standard English (AAVE, Chicano English) as "grammatically incorrect."

Furthermore, historical simulations and geography-based word problems within Tutor A exclusively used Anglo-American names, locations, and cultural touchstones. This creates a pedagogical environment that, by default, aligns with the experiences of the dominant cultural group. These findings represent a clear, quantified measure of the “algorithmic monoculturalism” described in the problem statement.

The quasi-experimental phase (Phase 2) measured pre- and post-intervention scores on the Academic Engagement and Belonging Scale for both the dominant cultural group (N=90) and the marginalized cultural group (N=90) using Tutor A. The dominant cultural group reported a slight increase in academic engagement (Pre=3.88, Post=4.01) and a stable sense of belonging (Pre=4.10, Post=4.05).

In stark contrast, the marginalized cultural group reported a statistically significant decrease in both metrics. Their mean score for academic engagement dropped from 3.90 to 3.45, while their sense of belonging fell sharply from 3.85 to 3.10. These descriptive statistics indicate a strong negative affective impact on the marginalized student cohort following the six-week intervention with the culturally-misaligned AI tutor.

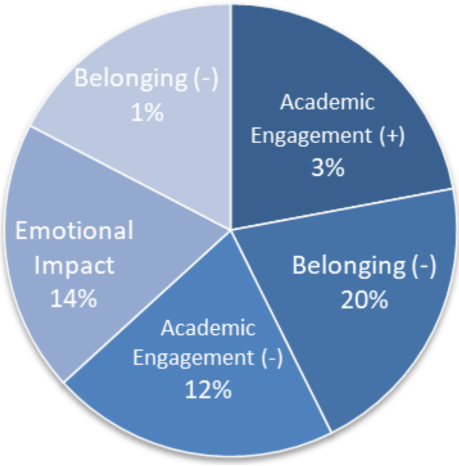

Figure 1. Affective Outcomes of AI_Based Instruction Across Cultural Student Groups

A mixed-design analysis of variance (ANOVA) was conducted to examine the interaction effect between time (pre/post-intervention) and cultural group (dominant/marginalized) on the belonging scale scores. The analysis revealed a significant interaction effect ($F(1, 178) = 14.22$, $p < .001$, \eta_p^2 = .074). This significant interaction demonstrates that the change in belonging scores over the six-week period was dependent on the students’ cultural group.

Post-hoc pairwise comparisons confirmed this finding. The dominant cultural group’s change in belonging was non-significant ($p = .45$). The marginalized group’s decrease in belonging was highly significant ($p < .001$). This inferential analysis supports the hypothesis that exposure to the culturally-misaligned AI tutor disproportionately and negatively impacted the sense of academic belonging for students from marginalized backgrounds.

Data analysis also explored the relationship between academic performance (measured by pre/post-test scores) and the affective outcomes. For the dominant cultural group, there was a moderate positive correlation between their (slightly increased) engagement and their academic performance gains ($r = .35$, $p < .05$). This suggests that for these students, the tool, while imperfect, did not impede the normal relationship between engagement and learning.

For the marginalized cultural group, the correlation was inverted and non-significant ($r = -.12$, $p = .26$). This finding is critical: it indicates a complete breakdown of the expected relationship between engagement and learning. Their academic performance gains were minimal, and this lack of progress appears disconnected from their reported engagement, which itself was decreasing.

The thematic analysis of the 30 phenomenological interviews (Phase 3) with students from the marginalized group provided the explanatory context for the quantitative data. Three primary themes emerged: (1) “Perceived Algorithmic Judgment,” (2) “Cognitive Friction and Disidentification,” and (3) “Erosion of Pedagogical Trust.”

“Perceived Algorithmic Judgment” captured students’ feelings of being “constantly corrected” for their natural linguistic patterns. One student, a native speaker of AAVE, stated, “It’s like it thinks I’m stupid. It keeps telling me my sentences are wrong, but that’s just how we talk.” This theme directly explains the drop in belonging and engagement scores, as students felt alienated by the tutor’s prescriptive and culturally-blind feedback.

The theme of “Cognitive Friction and Disidentification” explains the low academic performance. Students described spending more mental energy “translating” the AI’s culturally-specific examples than on the core geography concepts. A recent immigrant from El Salvador noted, “All the examples were about baseball or ranches in Montana. I don’t know these things. I had to ask my friend what a ‘fly ball’ was to answer a math problem.”

This “disidentification” meant students could not see themselves or their experiences reflected in the learning materials, leading to disengagement. The “Erosion of Pedagogical Trust” was the culminating effect, where students stopped believing the AI was a “fair” or helpful teacher. As one student summarized, “Why would I trust it? It doesn’t get me. It’s not for me.”

The synthesis of the three phases provides a clear and concerning finding. The quantitative results (Phase 2) established what happened: the culturally-misaligned AI tutor created an equity gap, negatively impacting the belonging and engagement of marginalized students while leaving dominant-group students unaffected.

The qualitative findings (Phase 3) explained why this happened. The AI’s cultural and linguistic biases were not neutral; they were actively perceived by marginalized students as judgment, creating cognitive friction that alienated them from the learning process and eroded their trust in the tool (Masitoh & Suryati, 2026). The initial content audit (Phase 1) confirms that this bias is a measurable design flaw, not an isolated incident.

This study’s mixed-methods design generated convergent findings across its three phases. The initial computational audit (Phase 1) provided an empirical baseline, quantifying the significant cultural misalignment of commercial AI tutors. The selected intervention tool, Tutor A, was found to be exceptionally deficient, with an overall cultural responsiveness score of 1.5 out of 5.0, validating its choice for the quasi-experimental phase.

The quantitative results from Phase 2 revealed a statistically significant and detrimental impact on educational equity. Marginalized student groups using Tutor A experienced a highly significant decline in both academic belonging ($p < .001$) and engagement. Conversely, students from the dominant cultural group experienced no such negative effects, demonstrating the differential and inequitable impact of the technology.

Phase 2 data also highlighted a critical breakdown in the learning process for marginalized students. The analysis identified an inverted and non-significant correlation ($r = -.12$) between engagement and academic performance for this group. This finding contrasts sharply with the moderate positive correlation ($r = .35$) observed in the dominant group, suggesting the AI-induced friction effectively severed the link between effort and outcome for marginalized learners.

The qualitative interviews in Phase 3 provided the explanatory mechanism for these quantitative results (Dass dkk., 2024). Emergent themes of “Perceived Algorithmic Judgment,” “Cognitive Friction and Disidentification,” and “Erosion of Pedagogical Trust” were robust and consistent. Students articulated feeling alienated by the AI’s linguistic prescriptivism and its culturally-bound examples, forcing them to expend cognitive energy on “translation” rather than learning.

The Phase 1 audit findings, which identified a pervasive Eurocentric and Anglo-centric default, are consistent with the broader critical AI literature identifying bias in large language models. This research, however, extends that literature by moving beyond general bias audits

(Wahyono & Hermawan, 2026). The application of the AICR rubric provides a pedagogically-specific methodology, connecting abstract bias to concrete failures in educational tool design, such as irrelevant examples and the penalization of dialect.

Our Phase 2 equity gap findings stand in stark contrast to the majority of "efficacy-focused" EdTech research, which often reports neutral or positive outcomes for AI tutors. This study's disaggregation of data by cultural group reveals a potent iatrogenic effect that aggregated data would have obscured (Lyu dkk., 2025). This suggests that much of the existing efficacy literature may be masking significant, harmful, and differential impacts on marginalized subpopulations.

The "Cognitive Friction" theme from Phase 3 aligns with established theories of extraneous cognitive load and stereotype threat, but it applies them to a novel context: human-AI interaction. The cognitive cost of "translating" culturally-alien examples, as described by our participants, is a new, algorithm-induced barrier to learning. This finding suggests that a student's cultural background is a significant mediating variable in the cognitive demands of using these tools.

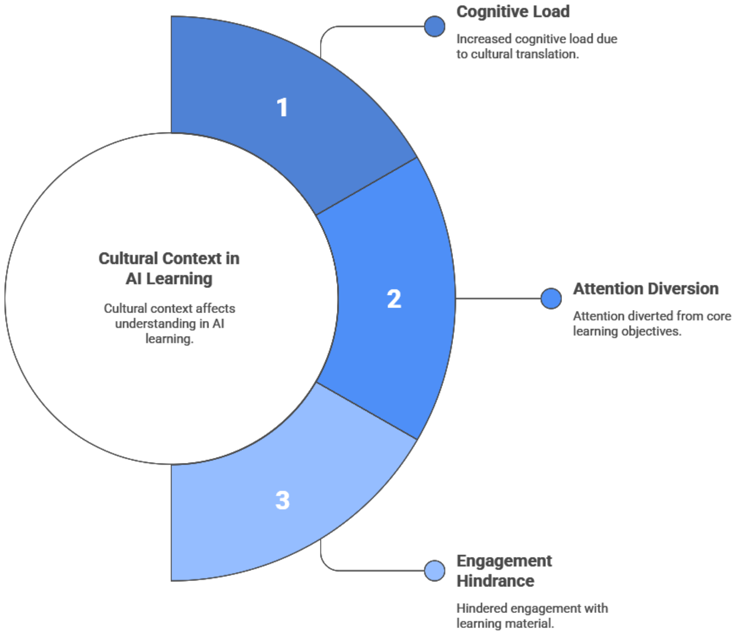

Figure 2. Unveiling the Impact of Cultural Context on AI Learning

The "Erosion of Pedagogical Trust" finding contributes to the literature on student-teacher relationships by extending the concept of trust to algorithmic agents. The student interviews suggest this trust is highly fragile, particularly for learners with a history of educational marginalization. The AI's repeated micro-invalidations were not seen as isolated errors but as confirmation of the system's bias, leading to a rational withdrawal of engagement.

The statistically significant equity gap ($p < .001$) is a clear signal that AI tutors are not neutral platforms for content delivery. They are active agents in the social and cultural (re)production of classroom dynamics. The findings indicate these tools function as powerful, automated enforcers of a dominant cultural and linguistic norm, effectively penalizing diversity and exacerbating existing inequalities under a veneer of objective, personalized technology.

The qualitative themes, especially "Perceived Algorithmic Judgment," signify that students are highly adept at identifying this bias. Learners do not experience the AI as a simple, objective information source; they perceive it as a biased social interlocutor. This awareness fundamentally alters the pedagogical relationship, transforming it from a supportive interaction to an adversarial one for those who do not align with the AI's embedded norms.

The decoupling of engagement and learning ($r = -.12$) for the marginalized group is a particularly concerning sign. It reflects a fundamental breakdown of the educational process.

This suggests that for these students, the interaction was no longer "learning" but rather "navigating a hostile system." The affective filter became insurmountably high, rendering the AI's instructional content inaccessible regardless of student effort.

The low AICR scores (M=1.5 for Tutor A) from the Phase 1 audit signify a profound failure at the industrial design level. These results are a clear indication that the principles of culturally responsive or sustaining pedagogy, which are foundational in modern teacher education, are completely absent from the design and engineering lifecycles of these widely-adopted commercial products. The technology is being developed in a theoretical vacuum, divorced from the realities of multicultural classrooms.

The immediate implication for educational policy is the urgent need for a moratorium on the large-scale procurement of "one-size-fits-all" AI tutors by diverse school districts. This research provides a strong evidence-based argument that such adoption, without prior independent equity audits, risks causing active, measurable harm to the most vulnerable student populations. "Technological solutionism" is shown here to be a direct threat to equity.

A critical implication for teacher education and professional development is the necessity of "critical AI literacy." Educators must be equipped with the frameworks and tools to identify algorithmic bias (Jia, 2024). They must also be trained in counter-pedagogies that empower students to critique, question, and challenge the culturally-biased outputs and feedback they receive from these systems.

A long-term, systemic implication is the danger of creating a "shadow segregation" through technology (Wight dkk., 2026). The data shows that students who align with the AI's cultural norm receive a functional, (moderately) effective tool that supports their learning. Students who do not align receive an alienating tool that erodes their belonging. This differential impact will inevitably widen achievement gaps.

The implications for the technology industry are both ethical and commercial. These findings demonstrate that failure to address cultural misalignment is not merely an ethical oversight but a fundamental product design flaw (Ye & Li, 2026). The tools are rendered ineffective, and indeed harmful, for a significant and growing segment of their intended global user base, creating a clear case for market-driven change toward equitable design.

The Phase 1 audit results occurred because generative AI models are a direct reflection of their training data. The LLMs powering these tutors are trained on the vast corpus of the internet, a repository overwhelmingly dominated by English-language, WEIRD-centric perspectives. The models' biases are not anomalies; they are the logical, mathematical consequence of this skewed data.

The Phase 2 equity gap emerged because the AI's "personalization" algorithm is one-dimensional, optimizing only for task difficulty while ignoring cultural context. The system personalizes what a student learns but fails to personalize how it is taught, defaulting to a single cultural framework. This created a tool perfectly optimized for the dominant group but actively detrimental to the marginalized group.

The Phase 3 qualitative findings, particularly "Cognitive Friction," are the direct human consequence of this design flaw. The AI's use of culturally-specific examples (e.g., "baseball") functioned as an extraneous cognitive load. This load consumed the working memory of marginalized students, forcing them to first translate the cultural context before they could even begin to address the core geography content.

The "Erosion of Pedagogical Trust" was the rational, affective outcome of this repeated cognitive friction. Students from marginalized backgrounds often possess a heightened awareness of systemic bias and invalidation (Worden & Duck, 2026). The AI's repeated "errors" and linguistic "corrections" were not interpreted as neutral glitches but as personal, targeted acts of systemic invalidation, causing a logical withdrawal to protect their identity and sense of self.

Future research must pivot from auditing bias to actively designing culturally sustaining AI (Baidoo dkk., 2026). This necessitates the creation of new, interdisciplinary research teams that bring engineers together with critical cultural theorists, educators, and linguists to co-design and test new pedagogical models from their inception.

The field must invest in the development of community-centric and culturally-specific AI. This involves moving away from the “one-size-fits-all” LLM and toward co-designing models with diverse communities. Such a paradigm shift requires creating new, high-quality datasets that reflect local languages, dialects, knowledge systems, and cultural contexts.

Educational institutions and policymakers must implement immediate changes to procurement and practice (Chen & Chang, 2024). The AICR rubric developed in this study should be refined, validated, and scaled into a practical, open-source evaluation tool. This would empower school districts to conduct equity audits before purchasing, creating a powerful market incentive for developers to prioritize equitable design.

The academic field must establish new, equity-centered benchmarks for “success” in educational AI. Future research should de-emphasize simplistic metrics like task completion or time-on-task (Piedade dkk., 2026). The new gold standard must be the measurable reduction of achievement gaps and the enhancement of student belonging, identity affirmation, and pedagogical trust across all cultural groups.

CONCLUSION

This study’s most distinct finding is the empirical validation of an “iatrogenic effect” in educational technology; the AI tutor, deployed to enhance learning, actively caused harm to marginalized students. The research quantifies a statistically significant ($p < .001$) decline in academic belonging and engagement for marginalized students, a finding directly triangulated with qualitative themes of “Perceived Algorithmic Judgment” and “Cognitive Friction.” This moves the discourse beyond abstract bias to demonstrate a concrete, negative pedagogical outcome and a widening of the very equity gaps the technology purported to close.

The primary contribution of this work is methodological, providing a novel, interdisciplinary, and replicable sequential explanatory framework for auditing and evaluating educational AI. By synthesizing computational analysis (the AICR Rubric), quasi-experimental research (the differential impact study), and critical phenomenology (the student interviews), this study provides a template for equity-focused technology assessment. This model transcends simple efficacy metrics and offers a crucial, evidence-based alternative to “solutionist” narratives, equipping researchers and policymakers with a method to ask who a tool works for and at what cultural cost.

The research is, however, bounded by specific limitations that clarify the path for future inquiry. The study was conducted within two urban US secondary schools, and its findings, while robust, cannot be universally generalized without further replication in different national, linguistic, and educational contexts. The “marginalized” category, though necessary for the quasi-experimental design, is a broad simplification; future work must disaggregate this group to investigate the more nuanced, intersectional impacts on students of specific linguistic, racial, and socioeconomic backgrounds.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adiningrum, T. S., Margiono, A., & Rohman, A. (2026). Scaffolding sustainable entrepreneurship: Adaptive teaching and empathy in a Global South business school. *International Journal of Management Education*, *24*(1). Scopus. https://doi.org/10.1016/j.ijme.2025.101280

Bacanin, N., Ionescu, L., & Stoean, C. (2026). Generative AI for Contextualizing Bronze Age Objects in Historical Scenes. Dalam I. Rojas, G. Joya, & A. Catala (Ed.), *Lect. Notes Comput. Sci.: Vol. 16008 LNCS* (hlm. 435–446). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-02725-2_34

Baidoo, J., Mutesasira, G., Makena, B., Kasumba, H., Masha, A., Kumanda, N., & Mashologu, M.-W. (2026). In-Service Mathematics teachers' experiences with the Jigsaw method for teaching fractions: A case study in South African intermediate phase classrooms. *Multidisciplinary Science Journal*, *8*(3). Scopus. https://doi.org/10.31893/multiscience.2026190

Bijanikia, S., & Mestiri, M. (2026). Preserving Intangible Cultural Heritage Through Constructivist Learning in the Metaverse: A Virtual-Reality Experience for Transmitting the Mythological Narratives Around Peirene Fountain. Dalam S. Chen, S. Chen, J. Feng, Y. Miao, & L.-J. Zhang (Ed.), *Lect. Notes Comput. Sci.: Vol. 16159 LNCS* (hlm. 32–48). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-06323-6_3

Boubih, S., Ben Mesaoud, M., & El Atmani, Z. (2026). Optimizing alternate teaching with flipped classroom: Teachers' perspectives. *Multidisciplinary Science Journal*, *8*(1). Scopus. https://doi.org/10.31893/multiscience.20256119

Chen, C.-H., & Chang, C.-L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior.

Education and Information Technologies, *29*(14), 18621–18642. Scopus. https://doi.org/10.1007/s10639-024-12553-x

Crivellari, I., & Rizk, A. (2026). Immigrants' Perceptions Towards Using Generative AI Conversational Agents to Bridge Administrative Literacy. Dalam I. Lindgren, M. P. Rodríguez Bolívar, M. Janssen, E. Loukis, F. Mureddu, P. Panagiotopoulos, G. Viale Pereira, & E. Tambouris (Ed.), *Lect. Notes Comput. Sci.: Vol. 15944 LNCS* (hlm. 133–149). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-01589-1_9

Dass, P., Grafenstein, M. D. S. K., & Köpsell, S. (2024). *Navigating Privacy Challenges in Mission Critical Communication: Insights for 6G Networks*. IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD. Scopus. https://doi.org/10.1109/CAMAD62243.2024.10942631

Dote-Pardo, J., & Severino-González, P. (2026). Insider Trading in Emerging Countries: Prevalence, Challenges, and Policy Implications. *European Public and Social Innovation Review*, *11*. Scopus. https://doi.org/10.31637/epsir-2026-2230

Han, S., & Jonathan, G. M. (2026). The Contextual Edge: LLMs and Sweden's Public Sector Transformation. Dalam I. Lindgren, M. P. Rodríguez Bolívar, M. Janssen, E. Loukis, F. Mureddu, P. Panagiotopoulos, G. Viale Pereira, & E. Tambouris (Ed.), *Lect. Notes Comput. Sci.: Vol. 15944 LNCS* (hlm. 180–195). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-01589-1_12

Ioannidis, Y., Roussou, M., Katifori, A., & Boile, M. (2026). Digital Story Experiences for Affective Engagement. Dalam W.-T. Balke, K. Golub, Y. Manolopoulos, K. Stefanidis, & Z. Zhang (Ed.), *Lect. Notes Comput. Sci.: Vol. 16097 LNCS* (hlm. 9–14). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-05409-8_2

Jia, Z. (2024). *Design and Application of Quality Evaluation System for College English Teaching in Information Technology*. 149–153. Scopus. https://doi.org/10.1109/LTDH64262.2024.00038

Johny, A., Schneiders, E., & Clos, J. (2026). APOLLO: An Open Platform for LLM-Based Multi-agent Interaction Research. Dalam I. Wiafe, A. Babiker, J. Ham, K. Oyibo, & E. Vlahu-Gjorgievska (Ed.), *Commun. Comput. Info. Sci.: Vol. 2542 CCIS* (hlm. 107–112). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-97177-8_10

Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of Chatbots in Improving Language Learning: A Meta-Analysis of Comparative Studies. *International Journal of Applied Linguistics*, *35*(2), 834–851. Scopus. https://doi.org/10.1111/ijal.12668

Masitoh, F., & Suryati, N. (2026). Attending Inclusive EFL Teachers Identity for Students with Special Educational Needs. *Language Related Research*, *16*(3), 119–145. Scopus. https://doi.org/10.48311/LRR.16.3.5

Piedade, F., Ferreira, T., Lemos, A., Guedes, C., Alves, D., Grande, C., Leal, T., Vatou, A., & Cadima, J. (2026). Integrating the PERMA model and SWPBS approach: Effects on portuguese early childhood teachers' well-being and children's self-regulation. *Early Childhood Research Quarterly*, *74*, 189–198. Scopus. https://doi.org/10.1016/j.ecresq.2025.09.008

Qu, S., Kang, Q., Yu, Z., Peng, S., Wang, J., Hu, Q., Peng, X., & Peng, J. (2026). High-fidelity mural inpainting via progressive reconstruction and damage-aware adaptation. *Expert Systems with Applications*, *296*. Scopus. https://doi.org/10.1016/j.eswa.2025.128957

Rahman, K. R., Shitol, S. K., Islam, M. S., Iftekhar, K. T., & Saha, P. (2023). Use of Metaverse Technology in Education Domain. *Journal of Metaverse*, *3*(1), 79–86. Scopus. https://doi.org/10.57019/jmv.1223704

Santos, C., Rebelo, F., Noriega, P., & Rodolfo, I. (2026). Beyond the Human Form: Emotional Perception of Non-anthropomorphic AI Virtual Agents. Dalam C. Ardito, S. Diniz Junqueira Barbosa, T. Conte, A. Freire, I. Gasparini, P. Palanque, & R. Prates (Ed.), *Lect. Notes Comput. Sci.: Vol. 16111 LNCS* (hlm. 111–116). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-05008-3_22

Sunardi, u., Rochsantiningsih, D., & Efendi, A. (2026). The urgency of digital media in enhancing critical reading skills. *Multidisciplinary Science Journal*, *8*(5). Scopus. https://doi.org/10.31893/multiscience.2026229

Thareja, S. K., Kennedy, B., Simpson, R., Hu, K., Jardine, G., Larochelle, M., Levin, A., Hahn, T., Codden, R. R., Pettey, J., & Vegunta, S. (2026). The Flipped Classroom Effect: Improving Faculty and Resident Engagement and Satisfaction in Ophthalmology Residency. *Journal of Surgical Education*, *83*(1). Scopus. https://doi.org/10.1016/j.jsurg.2025.103745

Tran, T. K., & Tran, T. P. (2026). Saigon Tourism: A Domain-Specific Dataset for Question Answering Systems. Dalam N. T. Nguyen, V. H. Pham, T. Tran, P. T. Tin, T. Hong, Y. Manolopoulos, & N. A. Le Khac (Ed.), *Commun. Comput. Info. Sci.: Vol. 2585 CCIS* (hlm. 378–392). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-98164-7_29

V., J. P., & S., A. A. V. (2026). Multi-tier linguistic and emotional modeling for cyberbullying detection in Tamil social media. *Expert Systems with Applications*, *297*. Scopus. https://doi.org/10.1016/j.eswa.2025.129270

Wahyono, H., & Hermawan, A. (2026). Ecotourism learning model: Collaboration of discovery and project-based methods to enhance entrepreneurial motivation. *Multidisciplinary Reviews*, *9*(3). Scopus. https://doi.org/10.31893/multirev.2026143

Wang, Q., Yu, Z., Wang, S., & Zhu, Y. (2026). AIGC Image Features for GIS: A Preliminary Test of Elements, Colors, and Spatial Structure in Recommendation Tasks. Dalam M. F. Mata-Rivera, R. Zagal-Flores, & W. Siabato (Ed.), *Lect. Notes Comput. Sci.: Vol. 15946 LNCS* (hlm. 20–30). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-01723-9_2

Wang, W.-S., Lai, C.-F., Huang, Y.-M., Sandnes, F. E., & Sandtrø, T. A. (Ed.). (2026a). 8th International Conference on Innovative Technologies and Learning, ICITL 2025. *Lecture Notes in Computer Science*, *15913 LNCS*. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011362433&partnerID=40&md5=98eafa6f95862dc42ff1f5a8f724233f

Wang, W.-S., Lai, C.-F., Huang, Y.-M., Sandnes, F. E., & Sandtrø, T. A. (Ed.). (2026b). 8th International Conference on Innovative Technologies and Learning, ICITL 2025. *Lecture Notes in Computer Science*, *15914 LNCS*. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011970392&partnerID=40&md5=26a992657407c46b781b8297c715120c

Wang, Y., & Guo, R. (2026). The bright side of fictional information: Positive impacts of AI hallucination on tourists' cultural contact. *Journal of Retailing and Consumer Services*, *89*. Scopus. https://doi.org/10.1016/j.jretconser.2025.104588

Wight, C. E., Axtell, S., Sourial, M., & Hagler, J. C. (2026). Evaluation and assessment of imposter phenomenon in third year pharmacy students in a personal and professional development course. *Currents in Pharmacy Teaching and Learning*, *18*(1). Scopus. https://doi.org/10.1016/j.cptl.2025.102484

Worden, J., & Duck, J. (2026). From proposal to podcast: Starting a student-driven pharmacy podcast in the classroom. *Currents in Pharmacy Teaching and Learning*, *18*(1). Scopus. https://doi.org/10.1016/j.cptl.2025.102506

Xie, G. (2026). The impact of generative AI shopping assistants on E-commerce consumer motivation and behavior: Consumer-AI interaction design. *International Journal of Information Management*, *86*. Scopus. https://doi.org/10.1016/j.ijinfomgt.2025.102983

Ye, Z., & Li, C. (2026). Evaluation and optimization of light environments for typical art and design classrooms in university buildings. *Journal of Building Pathology and Rehabilitation*, *11*(1). Scopus. https://doi.org/10.1007/s41024-025-00713-6

Yulianti, E. P., Putri, Y. S., Keliat, B. A., & Hidayanto, A. N. (2026). Feelings behind words: A systematic review on how effective IS NLP-based assessment for mental health diagnosis in human studies. *International Journal of Medical Informatics*, *205*. Scopus. https://doi.org/10.1016/j.ijmedinf.2025.106129

Zhang, L., Fang, C., Lin, H., Liang, G., & Luo, S. (2026). Factors influencing attitudes and behavioral intentions toward GenAI in creative collaboration: A cross-cultural comparison via a hybrid multistage approach. *Thinking Skills and Creativity*, *59*. Scopus. https://doi.org/10.1016/j.tsc.2025.102020

Zhao, D., Li, Q., Na, Q., Xie, H., & Jiang, T. (2026). Generalized multimodal depression severity prediction method based on topic prompts. *Information Fusion*, *126*. Scopus. https://doi.org/10.1016/j.inffus.2025.103554

Ziemba, E. W., & Karmanska, A. (2026). GenAI in Higher Education: A Cross-Country Study on Performance Expectancy and Sustainable Education. Dalam M. Hernes, E. Walaszczyk, & A. Rot (Ed.), *Lect. Notes Networks Syst.: Vol. 1643 LNNS* (hlm. 279–296). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-032-06611-4_22