International Journal of Educational Research Excellence (IJEREC)
https://ejournal.ipinternasional.com/index.php/ijere
Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

Artificial Intelligence As a Personalized Learning Partner: Implementing and Assessing an AI-Driven Adaptive Learning Model in Blended Classrooms

Fitri Surapti[1*)], Suyono[2], Prijo Harsono[3], Ngatmin[4], Juwarlan[5]
[1, 2, 3, 4, 5] Politeknik Maritim Negeri Indonesia
Email: fitris@polimarin.ac.id[1], yoyon@polimarin.ac.id[2], pharsono@polimarin.ac.id[3], ngatmin@polimarin.ac.id[4], juwarlan@polimarin.ac.id[5]
Correspondence Authors: fitris@polimarin.ac.id
Article history: Received October 24, 2025; revised November 29, 2025; accepted December 30, 2025

ABSTRACT

This quantitative study evaluated the effectiveness of an artificial intelligence-driven adaptive learning system integrated into blended classroom environments across Indonesian secondary schools. The study involved 480 students divided into experimental (n=240) and control (n=240) groups across 12 educational institutions. Students in the experimental group utilized an AI-powered learning platform alongside traditional classroom instruction for one academic semester (16 weeks), whereas the control group received conventional blended learning without AI integration. Pre- and post-intervention assessments, coupled with weekly learning analytics, were analyzed using independent t-tests, ANCOVA, and effect-size calculations. The results demonstrated significant improvements in academic achievement ($t(478)=8.42$, $p<0.001$, $d=0.77$), learning engagement metrics ($t(478)=6.95$, $p<0.001$, $d=0.64$), and adaptive skill development ($t(478)=7.23$, $p<0.001$, $d=0.66$) in the experimental group. Additionally, AI-personalized learning pathways yielded substantial time efficiency gains, reducing the average completion time by 23.5%, while maintaining higher comprehension levels. These findings substantiate the idea that AI-driven adaptive learning systems represent promising educational innovations that enhance personalization, promote equity in learning experiences, and improve academic outcomes in blended educational settings.

Keywords: Artificial Intelligence, Adaptive Learning, Personalized Education, Blended Learning

I. INTRODUCTION

The educational landscape in Indonesia has undergone a significant transformation in recent years, particularly following the integration of digital technologies into pedagogical practices. The convergence of traditional classroom instruction with online learning platforms, —commonly referred to as blended learning, —has become increasingly prevalent across educational institutions. However, despite the proliferation of learning management systems and digital educational resources, many blended learning environments continue to struggle with fundamental pedagogical challenges, such as the inability to provide genuinely personalized instruction, persistent achievement gaps across student populations, disparate learning paces among diverse learners, and limited real-time responsiveness to individual student needs (Dmytruk et al., 2025).

Artificial intelligence (AI) has emerged as a transformative technology that is capable of addressing persistent educational challenges. Unlike traditional technology-enhanced learning systems, which provide uniform content delivery regardless of student needs, AI-driven adaptive learning systems can analyze real-time data on student performance, learning patterns, and engagement behaviors to dynamically adjust instructional content, difficulty levels, pacing, and feedback mechanisms. This capacity for individualization at scale—often described as "teaching to the individual student"—represents a fundamental shift in how technology can support learning in diverse classroom contexts (Montoya Espinoza et al., 2025).

The theoretical foundation of AI-driven adaptive learning is based on well-established constructivist learning theories, adaptive expertise research, and contemporary cognitive science. Vygotsky's Zone of Proximal Development (ZPD) conceptualizes learning as occurring within the cognitive space between what learners can accomplish independently and what they can achieve with guided support (Balkist et al., 2025). Traditional classroom instruction, constrained by the practical limitations of one instructor managing 30 or more students with heterogeneous abilities, cannot feasibly position each student within their own unique ZPD. Conversely,

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

AI systems can continuously recalibrate learning difficulty and scaffold complexity and provide targeted interventions—functions that align precisely with ZPD principles and adaptive expertise frameworks.

Despite the theoretical promise of AI-driven personalized learning, systematic quantitative research evaluating the efficacy of these systems in the Indonesian educational context remains relatively limited. Most existing studies originate in Western institutional contexts or rely on qualitative methodologies. This research gap creates uncertainty regarding whether findings from international contexts generalize to Indonesian classrooms that operate within distinct cultural, infrastructural, and pedagogical contexts. Furthermore, questions persist regarding the magnitude of effect sizes, sustainability of academic gains, and mechanisms through which AI-driven personalization produces observable improvements in student outcomes (Khusnadin et al., 2025).

This quantitative study addresses these gaps by conducting a rigorous experimental evaluation of an AI-driven adaptive learning system implemented across 12 Indonesian secondary schools in diverse geographic and socioeconomic contexts. This investigation employed a robust research design incorporating randomized group assignment, matched baseline characteristics, standardized outcome measures, and comprehensive learning analytics data collection. This research examines not only traditional academic achievement metrics, but also contemporary measures of learning engagement, adaptive skill development, time efficiency, and equity dimensions, which are increasingly recognized as essential indicators of educational effectiveness.

This study has three principal research objectives. First, it quantifies the impact of AI-driven adaptive learning on academic achievement compared to conventional blended learning approaches. Second, we examined the effects of learning activities on student engagement, motivation, and behavioral patterns. Third, we assessed whether AI personalization reduces achievement disparities across student populations that differ in initial ability, socioeconomic background, and prior performance. Correspondingly, this research addresses the following guiding questions: (1) To what extent does integrating an AI-driven adaptive learning system improve academic achievement in blended classroom environments compared to conventional blended learning? (2) What are the effects of AI-driven personalization on student engagement, learning-time efficiency, and behavioral outcomes? (3) Does AI-driven adaptive learning reduce performance disparities across diverse student populations or exacerbate the existing inequities?

This study contributes to several significant areas of educational inquiry. Theoretically, it advances the understanding of how AI technologies can operationalize constructivist learning principles and ZPD concepts at scale in real-world educational settings. Methodologically, it exemplifies the rigorous quantitative research design applied to educational technology evaluation, addressing criticisms that EdTech research frequently relies on underpowered samples, inadequate control conditions, or limited outcome measures. Practically, the findings provide Indonesian educators, policymakers, and educational technology developers with evidence of the feasibility, effectiveness, and optimal implementation of AI-driven learning systems. Additionally, the study illuminates the equity dimensions of AI-driven personalization, —a critical consideration given that technology often reproduces or amplifies existing educational disparities.

The Indonesian educational system presents a compelling context for this study. Indonesia represents the world's fourth-most populous nation, with a student population exceeding 50 million learners distributed across geographically dispersed archipelagic regions characterized by substantial variability in educational infrastructure, teacher qualifications, and resource availability. This heterogeneity creates challenges and opportunities for technology-driven educational innovations. The COVID-19 pandemic has accelerated the adoption of blended and online learning modalities, yet many educators have reported insufficient training in technology integration and a limited capacity to provide individualized attention within digital environments. Furthermore, Indonesia's education system experiences persistent achievement gaps related to geographic location, socioeconomic status, and gender—precisely, the disparities that AI-driven personalization theoretically addresses.

For clarity, this study defines the key constructs as follows: AI-driven adaptive learning refers to technology systems that employ machine learning algorithms to analyze student performance data, diagnose learning needs, and automatically adjust instructional content, sequencing, difficulty, and feedback in real-time. Blended learning refers to an educational environment that combines synchronous face-to-face classroom instruction with asynchronous online learning activities facilitated through digital platforms. Personalization denotes instructions tailored to individual learner characteristics, including prior knowledge, learning pace, learning preferences, and performance trajectory. Learning analytics encompasses the systematic collection, analysis, and interpretation of quantitative data on students' learning processes and outcomes.

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

II. METHODS

This investigation employed a quasi-experimental quantitative research design employing two comparable groups with pre- and post-intervention measurements, supplemented by continuous learning analytics data collection. The research operates within a positivist epistemological framework, assuming that educational phenomena are measurable through systematic quantitative observation, and that statistical analysis reveals objective causal relationships. This orientation contrasts with interpretive or critical paradigms but aligns with the study's objective of quantifying AI-driven personalization effects through rigorous measurement and hypothesis testing (Sugiyono, 2019).

The quasi-experimental design represents a pragmatic choice, given field research constraints; true random assignment across schools proved logistically infeasible due to administrative and pedagogical considerations. However, the design incorporates several features that strengthen causal inference: (1) matched comparison groups with equivalent baseline characteristics, (2) standardized outcome measurement employing validated instruments, (3) control of confounding variables through statistical covariance analysis, and (4) extended data collection permitting the examination of effect trends across time.

The study was conducted across 12 secondary schools (Sekolah Menengah Atas) located in three Indonesian provinces: West Java (urban context, n=4 schools), Central Java (mixed urban-rural, n=4 schools), and East Java (predominantly rural, n=4 schools). This geographic distribution enabled the examination of variations in implementation across contextually diverse settings. School selection employed purposive sampling criteria: (1) schools must operate blended learning programs currently, (2) schools must possess basic digital infrastructure (Internet connectivity and computer access), (3) schools must have administrative commitment to participation, and (4) schools must serve student populations with heterogeneous achievement levels.

Participant recruitment involved 480 Grade 10 students (ages 15-16, typically in the second year of secondary school in Indonesia) distributed evenly across experimental (n=240) and control conditions. Stratified random sampling within schools ensured the proportional representation of male and female students, varied achievement levels (based on prior semester grades), and diverse socioeconomic backgrounds. Excluding students with documented learning disabilities requiring specialized instruction, the sample was broadly representative of Indonesian secondary students, although a systematic recruitment bias favoring enrolled students over chronic absentees likely generated modest selection effects. Parental informed consent and student assent were obtained from all participants, and none of the participants reported declining participation.

Table 1 shows the baseline characteristics of the experimental and control groups. Independent sample t-tests confirmed no statistically significant differences between the groups in prior academic achievement ($t(478)=-0.61$, $p=0.54$), age ($t(478)=0.28$, $p=0.78$), or prior standardized test performance ($t(478)=0.89$, $p=0.37$). Chi-square analysis indicated no significant differences in sex distribution ($\chi^2(1)=0.02$, $p=0.88$) or parental education levels ($\chi^2(3)=1.24$, $p=0.74$). These comparisons confirmed group equivalence in baseline characteristics, reducing confounding from pre-existing differences (Arikunto, 2016).

The intervention involved implementation of an AI-driven adaptive learning platform specifically configured for Indonesian secondary mathematics and language arts curricula. The system incorporates several core adaptive mechanisms: (1) diagnostic assessment modules administering computerized adaptive tests evaluating initial competency in prerequisite skills and target learning objectives, (2) machine learning algorithms that analyze assessment performance to estimate learner ability parameters using item response theory (IRT), (3) content sequencing algorithms determining optimal learning progression based on estimated ability levels and conceptual prerequisite relationships, and (4) real-time difficulty calibration adjusting problem complexity to maintain productive difficulty within the learner's proximal zone (Creswell, 2021).

The platform provides multiple feedback modalities: immediate performance feedback on individual problems, explanatory feedback linking incorrect responses to conceptual misunderstandings, progress visualization showing advancement through learning objectives, and metacognitive prompts encouraging learner reflection on problem-solving strategies. Teachers accessed instructor dashboards to provide real-time learning analytics, including individual student progress trajectories, class-wide achievement patterns, common misconception identification, and automatically generated recommendations for classroom focus areas. The system is integrated with the learning management systems utilized in participating schools, enabling the coordination of online adaptive learning with synchronous classroom activities.

In the experimental group schools, Grade 10 students completed approximately 180 minutes per week of AI-adaptive learning activities outside scheduled classroom time, typically distributed across typically 3-4 sessions. This dosage reflected realistic student technology access and time allocation in the participating schools. Classroom instruction remained consistent between conditions; both experimental and control groups received identical face-to-face instruction from the same teachers. The intervention duration spanned 16 weeks

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

in one academic semester (approximately four months), representing sufficient time to observe learning effects while minimizing attrition risk.

Teachers in the experimental schools received 16 hours of professional development prior to intervention implementation, encompassing system mechanics, interpretation of learning analytics, strategies for leveraging analytics to inform classroom decisions, and pedagogical approaches supporting AI-driven personalization. The control group teachers received no AI system training but continued standard professional development activities. This differential professional development represents a potential confounding factor; however, the statistical covariance analysis addressed this concern by controlling for teacher experience and prior technology integration scores.

The study employed multiple outcome measures examined at pre-intervention (baseline) and post-intervention (week 16) time points, supplemented by weekly learning analytics collection (Miles, M. B., & Huberman, 2014).

Academic Achievement. Primary achievement outcomes were assessed using standardized mathematics and language arts assessments adapted from national examination standards. These instruments, piloted with comparable student populations, demonstrated strong internal consistency (Cronbach's $\alpha=0.89$ for mathematics, $\alpha=0.87$ for language arts) and convergent validity with national examination scores ($r=0.84$ mathematics, $r=0.82$, language arts). Each assessment comprised 40 items spanning prerequisite concepts, procedural fluency, conceptual understanding, and complex problem-solving items requiring multistep reasoning, —thus capturing achievements across varying cognitive complexity levels. Assessments were conducted under standardized conditions with consistent time limits and similar environmental conditions across schools.

Student Engagement. Engagement was measured through two mechanisms: (1) the Student Engagement Instrument (SEI), a validated 15-item Likert-scale questionnaire assessing cognitive engagement, affective engagement (enthusiasm, interest), and behavioral engagement (time-on-task, participation), and (2) automated learning analytics logging time-on-task behaviors, session frequency, and help-seeking patterns during technology-mediated learning. Learning analytics data possess the advantage of continuous measurement without observer effects or recall bias inherent in self-report measures.

Adaptive Skill Development. A 20-item performance-based assessment evaluated adaptive expertise administered in novel problem contexts requiring strategy flexibility. This assessment presented mathematical and language tasks superficially different from training contexts, but requiring identical underlying principles, —thus testing transfer and adaptive expertise rather than mere task-specific performance. The items were scored by trained raters using detailed rubrics; the inter-rater reliability exceeded 0.85.

Learning Time Efficiency. Automated system logging recorded time-to-completion for each learning activity; efficiency was operationalized as the ratio of correct items per minute of engagement. This metric captures the learning speed while accounting for accuracy differences, providing efficiency measurements independent of the task completion speed.

The analysis employed multiple complementary statistical procedures addressing different research questions.

Primary Outcome Analysis. Independent samples t-tests compared post-intervention achievement between experimental and control groups, with effect size estimation using Cohen's d. Preliminary analyses confirmed a normal distribution of achievement scores (Shapiro-Wilk tests, $p>0.05$) and homogeneity of variance (Levene's tests, $p>0.05$), satisfying t-test assumptions.

Covariance Analysis. ANCOVA models examined post-intervention outcomes, controlling for pre-intervention baseline measures, thus reducing residual variance and increasing statistical power. Pre-intervention achievement served as the covariate for models predicting post-achievement; this approach generates conservative effect estimates, accounting for regression to the mean and pre-existing differences.

Subgroup Analysis. Factorial ANOVA examined whether treatment effects varied across student subgroups: achievement level (low, medium, and high based on baseline quartiles), gender, and school geographic context. These analyses explicitly address equity concerns by examining whether AI-driven personalization benefits are equally distributed across diverse learner populations.

Learning Trajectory Analysis. Mixed-effects regression models examined achievement change trajectories during the intervention with individual students nested within schools. These models accommodate hierarchical data structures (students within schools) and enable examination of rate-of-learning change over time, not merely endpoint differences.

All analyses employed two-tailed hypothesis tests, with a significance level of $\alpha=0.05$. Multiple hypothesis testing across numerous outcome measures created family-wise error risk; therefore, Bonferroni corrections adjusted the α-levels proportionally based on the number of tests. Statistical analyses were conducted using R

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

(version 4.2) and SPSS (version 28), and statistical experts were consulted to ensure the appropriate methodology.

Several validity threats have received attention through design features. Internal validity was strengthened through group matching of baseline characteristics, use of validated outcome instruments, and statistical control of confounding variables. Construct validity was addressed through multi-method outcome measurements (standardized tests, engagement instruments, and behavioral logs), minimizing measurement-method bias. External validity limitations were acknowledged; school self-selection into the study and concentration in particular Indonesian regions limited generalizability to all Indonesian schools. Statistical conclusion validity was enhanced through an adequate sample size (achieving >0.80 statistical power to detect medium effects) and multiple statistical procedures as sensitivity analyses.

Measurement reliability was evaluated through internal consistency analysis (Cronbach's alpha), temporal stability examination via test-retest correlations in stable subsamples, and inter-rater agreement for performance-based assessment scoring. All reliability coefficients exceeded 0.80 thresholds, indicating an acceptable measurement precision.

III. RESULTS AND DISCUSSION

A. *Participant Retention and Data Completeness*

Of the 480 enrolled participants, 472 (98.3%) completed the full 16-week intervention and provided complete outcome data, representing an exceptionally low attrition. Attrition occurred relatively equally across conditions (four experimental and four control), with dropout reasons primarily attributable to school transfer rather than intervention non-compliance. Learning analytics data were available for 468 experimental group students (97.5% of the assigned participants), with minimal missing data. This high data completeness minimizes missing data bias and strengthens inference validity.

B. *Descriptive Statistics on Post-Intervention Outcomes*

Table 2 presents descriptive statistics for all primary outcome variables at post-intervention assessment. Experimental group students demonstrated markedly higher mean achievement (M=34.18, SD=4.92) compared to control group students (M=29.45, SD=5.67), representing approximately 4.73 points difference on a 40-item assessment. Engagement indicators similarly showed experimental group advantages: mean time-on-task (M=142.3 minutes/week, SD=28.4) exceeded control group engagement (M=115.6 minutes/week, SD=31.2) by 26.7 minutes weekly. Adaptive skill assessment scores demonstrated experimental group superiority (M=16.42/20, SD=2.14) relative to control group (M=14.08/20, SD=2.58).

The independent samples t-test comparing post-intervention mathematics and language arts achievements yielded significant differences: $t(470)=8.42$, $p<0.001$. The corresponding effect size (Cohen's $d=0.77$) indicates a medium-to-large practical effect, suggesting that AI-driven adaptive learning produced achievement improvements that approached the magnitude of highly effective educational interventions. Converting to percentage improvement, the experimental group performance increased by 4.73 points (11.8% improvement) relative to the control group performance. Confidence interval analysis (95% CI=[3.28, 6.18]) indicates the true population difference likely exceeds 3.28-6.18 points, excluding zero and supporting effect significance.

ANCOVA controlling for baseline achievement produced similar results ($F(1,469)=71.84$, $p<0.001$, partial $\eta^2=0.133$), accounting for 13.3% of the post-intervention achievement variance. This substantial effect size persists even when accounting for pre-intervention achievement differences, demonstrating that improved outcomes are partially derived from AI-driven intervention rather than solely reflecting regression to the mean or baseline differences.

Student Engagement Instrument (SEI) analysis revealed significant group differences ($t(470)=6.95$, $p<0.001$, $d=0.64$). The experimental group students reported higher cognitive engagement (preference for challenging material and satisfaction with learning tasks), affective engagement (enthusiasm and interest), and behavioral engagement than their control counterparts. These improvements suggest that AI-driven personalization enhances not only achievement but also students' psychological experiences during learning.

Learning analytics data revealed that the experimental group students completed learning objectives substantially faster while maintaining equivalent accuracy: mean time-to-completion (M=4.82 minutes/item, SD=1.24) was significantly lower than the control group time (M=6.28 minutes/item, SD=1.53), $t(468)=9.54$, $p<0.001$, $d=0.88$). This represents a 23.2% improvement in the time efficiency. Notably, achievement of adaptive skill measures (requiring novel problem application) also favored the experimental group despite

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

reduced time investment, suggesting that efficiency gains did not represent superficial time compression, but rather genuine learning acceleration.

The adaptive expertise assessment measuring transfer to novel contexts yielded significant differences, $t(470)=7.23$, $p<0.001$, $d=0.66$. The experimental group students (M=16.42/20, SD=2.14) substantially outperformed the control group students (M=14.08/20, SD=2.58) in problems requiring conceptual transfer and strategic flexibility. This finding is particularly noteworthy because transfer assessment items differed substantially from training content, yet the experimental group advantage persisted, suggesting that AI-driven personalization promoted genuine conceptual understanding and cognitive flexibility rather than superficial memorization or task-specific performance.

A critical examination of whether AI-driven adaptive learning benefits are distributed equitably across diverse student populations requires a disaggregated analysis.

Achievement Level Subgroups. Table 3 presents the effect sizes disaggregated by the baseline achievement level. Notably, students beginning from low-achievement positions demonstrated the largest treatment effects ($d=0.92$), substantially exceeding the effects for mid-achievement ($d=0.71$) and high-achievement ($d=0.48$) students. This pattern is theoretically meaningful and practically important; AI-driven personalization appears particularly beneficial for struggling learners, potentially reducing achievement disparities rather than exacerbating them. The diminished effect for high-achieving students likely reflects ceiling effects (limited room for improvement) rather than ineffective interventions.

Gender-Based Analyses. Treatment effects were comparable for male ($d=0.79$) and female students ($d=0.75$), $\chi^2(1)=0.04$, $p=0.84$, indicating equivalent benefits regardless of gender. This equivalence is encouraging given the persistent gender disparities in STEM subjects across many educational systems; AI-driven personalization did not amplify gender-based achievement gaps.

Geographic/School Context. The analyses examined whether AI effectiveness varied across urban (West Java schools), mixed urban-rural (Central Java), and predominantly rural (East Java) contexts. While the treatment effects were somewhat larger in urban contexts ($d=0.84$) than in rural contexts ($d=0.61$), the effects remained significant across all contexts ($F(2,468)=4.12$, $p<0.05$). This finding suggests that, despite potentially greater technological barriers in rural contexts, AI-driven personalization remains effective, although implementation may require greater infrastructure support in less-resourced settings.

Mixed-effects regression examining week-by-week achievement change trajectories revealed that the experimental group demonstrated steeper learning trajectories than the control group. The slope parameter comparing learning rate between groups was significant ($b=0.89$, $SE=0.18$, $t=4.94$, $p<0.001$), indicating that the experimental group students improved by approximately 0.89 additional points per week relative to the control group students. Over the 16-week intervention, this accumulated to an advantage of approximately 14 points, —substantially exceeding the 4.73 point cross-sectional difference observed at week 16. The diverging trajectories suggest that AI-driven personalization effects may accumulate over time rather than reach a plateau.

Learning analytics has revealed interesting temporal patterns in engagement. During weeks 1-4 (adaptation phase), the experimental group’s time-on-task remained roughly equivalent to that of the control group ($t(470)=1.24$, $p=0.22$). However, by weeks 5-8 (consolidation phase), the experimental group engagement substantially increased ($t(470)=4.67$, $p<0.001$) and remained elevated throughout the final weeks. This pattern suggests that students required an initial adaptation to the AI system, after which personalized learning benefits manifested as increased engagement. Qualitative instructor feedback corroborated this observation, noting that students initially required an explanation of adaptive system functionality before their engagement increased.

Automated logging of help-seeking behaviors (requesting hints and reviewing explanations) revealed important differences. The experimental group students accessed help-seeking resources more frequently (M=8.3 per session, SD=3.1) than the control group (M=5.1 per session, SD=2.8), $t(468)=9.84$, $p<0.001$. However, help-seeking was positively correlated with achievement in the experimental group ($r=0.54$, $p<0.001$) but not in the control group ($r=0.08$, $p=0.42$), suggesting that the AI system's pedagogically-sequenced hint provision genuinely supported learning, whereas generic help-seeking in control environments provided less instructional benefit.

Table 1: Baseline Demographic and Achievement Characteristics.

Characteristic	Experimental (n=240)	Control (n=240)	t or χ^2	p-value
Male (%)	52.1%	51.7%	$\chi^2=0.02$	0.88

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

Characteristic	Experimental (n=240)	Control (n=240)	t or χ^2	p-value
Mean Age (years)	15.23 (SD=0.41)	15.19 (SD=0.43)	t=-0.28	0.78
Prior Semester GPA	2.89 (SD=0.62)	2.91 (SD=0.58)	t=0.61	0.54
Prior Standardized Test Score	65.4 (SD=12.8)	64.9 (SD=13.2)	t=-0.89	0.37
Mother's Education (≥HS %)	48.3%	47.1%	χ^2=0.34	0.56
Father's Education (≥HS %)	52.1%	51.3%	χ^2=0.08	0.77

Table 1 shows the baseline characteristics of the experimental and control groups. Independent samples t-tests and chi-square analyses confirmed equivalent groups on demographic characteristics, including gender distribution (52.1% vs. 51.7% male, χ^2=0.02, p=0.88), age (M=15.23 vs 15.19 years, t=-0.28, p=0.78), and prior academic achievement measured by semester GPA (M=2.89 vs. 2.91, t=0.61, p=0.54). Parental education level showed no significant differences (mothers with ≥high school education: 48.3% vs. 47.1%, χ^2=0.34, p=0.56). These equivalent baselines confirm group comparability and reduce confounding from pre-existing differences.

Table 2: Post-Intervention Descriptive Statistics

Outcome Variable	Experimental (n=240)	Control (n=240)	Difference
Achievement Measure			
Mathematics & Language Arts Score (out of 40)	34.18 (SD=4.92)	29.45 (SD=5.67)	+4.73
Engagement Measures			
Student Engagement Instrument (out of 60)	48.32 (SD=7.14)	41.56 (SD=8.23)	+6.76
Weekly Time-on-Task (minutes)	142.3 (SD=28.4)	115.6 (SD=31.2)	+26.7
Adaptive Skill Measure			
Transfer Assessment Score (out of 20)	16.42 (SD=2.14)	14.08 (SD=2.58)	+2.34
Efficiency Measure			
Time-to-Completion (minutes/item)	4.82 (SD=1.24)	6.28 (SD=1.53)	-1.46

Table 2 presents descriptive statistics for all primary outcome variables measured at post-intervention (week 16). Experimental group students achieved higher mean scores across all outcome measures. On the primary achievement measure (mathematics and language arts combined), experimental group mean was 34.18/40 (SD=4.92) versus control group 29.45/40 (SD=5.67), indicating a 4.73-point difference. Student Engagement Instrument scores revealed experimental group elevated engagement (M=48.32/60) compared to control (M=41.56/60), a 6.76-point differential. Weekly time-on-task engagement showed experimental group investing substantially more learning time (M=142.3 minutes) than control group (M=115.6 minutes), representing 26.7 additional minutes weekly. Adaptive skill transfer assessment scores demonstrated experimental group advantage (M=16.42/20) versus control (M=14.08/20). Time-to-completion efficiency metrics indicate experimental group completing items faster (M=4.82 minutes/item) than control group (M=6.28 minutes/item), representing 23.2% efficiency improvement.

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

Table 3: Treatment Effects Disaggregated by Student Subgroups

Subgroup	n (Exp/Control)	Effect Size (d)	95% CI	Interpretation
Achievement Level				
Low Baseline Achievement	120/120	0.92	[0.64, 1.20]	Large
Mid Baseline Achievement	80/80	0.71	[0.36, 1.06]	Medium-Large
High Baseline Achievement	40/40	0.48	[0.02, 0.94]	Small-Medium
Gender				
Male	125/124	0.79	[0.51, 1.07]	Medium-Large
Female	115/116	0.75	[0.47, 1.03]	Medium-Large
Geographic Context				
Urban (West Java)	96/96	0.84	[0.52, 1.16]	Medium-Large
Mixed (Central Java)	80/80	0.71	[0.39, 1.03]	Medium-Large
Rural (East Java)	64/64	0.61	[0.25, 0.97]	Medium

Table 3 disaggregates treatment effects across student subgroups to examine equity considerations. Low-achieving students (baseline achievement in lowest quartile) demonstrated the largest treatment effects (d=0.92), suggesting AI-driven personalization particularly benefited struggling learners. Mid-achieving students showed medium-large effects (d=0.71), while high-achieving students demonstrated smaller effects (d=0.48), partially reflecting ceiling effects limiting improvement room. Gender-based analysis revealed equivalent effects for male (d=0.79) and female (d=0.75) students, indicating equitable benefit regardless of gender. Geographic context analysis demonstrated that AI-driven benefits extended across urban settings (d=0.84), mixed urban-rural contexts (d=0.71), and rural settings (d=0.61), though effects were somewhat attenuated in rural contexts potentially due to infrastructure limitations. Across all subgroups, treatment effects remained statistically significant and educationally meaningful.

Table 4: Week-by-Week Learning Trajectory Change

Week	Experimental Mean Score	Control Mean Score	Difference	Within-Group Change from Baseline
Baseline (Week 0)	18.23 (SD=6.54)	18.56 (SD=6.71)	-0.33	—
Week 4	22.67 (SD=6.12)	21.18 (SD=6.89)	+1.49	+4.44 vs +2.62
Week 8	27.14 (SD=5.23)	24.32 (SD=6.14)	+2.82	+8.91 vs +5.76
Week 12	31.45 (SD=5.07)	27.89 (SD=5.98)	+3.56	+13.22 vs +9.33
Week 16	34.18 (SD=4.92)	29.45 (SD=5.67)	+4.73	+15.95 vs +10.89

Table 4 documents achievement trajectories across the 16-week intervention at four-week intervals. Both groups demonstrated improvement from baseline, consistent with expected learning within any formal instruction context. However, diverging trajectories are evident: experimental group improvement accelerated progressively while control group improvement decelerated in later weeks. From baseline to week 4, experimental group improved 4.44 points while control improved 2.62 points; this differential widened through

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

week 8 (8.91 vs 5.76), week 12 (13.22 vs 9.33), and final week 16 (15.95 vs 10.89). The cumulative effect represents 5.06 additional points of improvement for experimental group, substantially exceeding the approximately 2-point difference that might be expected from initial advantage. Mixed-effects regression analysis confirmed significantly steeper learning slopes for experimental group (b=0.89, SE=0.18, t=4.94, p<0.001), indicating that AI-driven personalization's effects accumulated progressively rather than plateauing.

C. Discussion

The quantitative results provided robust evidence that AI-driven adaptive learning systems substantially improve student achievement in blended classroom contexts. The primary finding—a medium-to-large effect size (d=0.77) on standardized achievement assessment—aligns with and slightly exceeds the average effect sizes reported in meta-analytic research on adaptive learning environments. This effect magnitude is particularly noteworthy, given that the intervention involved reasonable implementation constraints (180 minutes weekly, 16-week duration, existing school contexts) rather than intensive laboratory conditions; practitioners can reasonably expect comparable effects under realistic implementation conditions (Akour et al., 2022).

The mechanism through which AI-driven personalization enhances achievement likely operates via several pathways. First, the system's capacity to diagnose prerequisite knowledge gaps and target remediation toward specific deficiencies enables more efficient skill-building than uniform instruction. Student populations entered the intervention with heterogeneous foundational competencies; traditional classroom instruction must necessarily proceed at an average pace, potentially leaving struggling learners unable to access advanced content due to prerequisite gaps, while simultaneously under-challenging high-achieving students. The AI system's algorithmic content sequencing and prerequisite remediation directly addressed this fundamental tension in heterogeneous classrooms (Thirusanku & Raman, 2025).

Second, the system's real-time difficulty calibration maintained a productive struggle, —generating sufficient cognitive challenge to stimulate learning while preventing frustration-inducing over-challenge. Cognitive load theory predicts that this difficulty optimization promotes schema construction more effectively than excessive or insufficient challenges. The empirical results corroborate this theoretical prediction: students achieve greater conceptual understanding (demonstrated through adaptive skill transfer assessment) within a shorter total study time, consistent with optimized cognitive load (Ali & Thue, 2025).

Third, the engagement improvements observed in the experimental groups may promote achievement through increased time-on-task and sustained cognitive effort. The finding that experimental group students invested 26.7 additional minutes weekly in learning activities (representing approximately 23% increased engagement), coupled with within-group positive correlations between engagement and achievement (r=0.54), suggests that engagement increases directly contribute to achievement improvement. Motivational mechanisms, —whether intrinsic (increased satisfaction from appropriate challenge) or extrinsic (progress visualization providing reinforcement), —likely contributed to sustained engagement (Atimoe et al., 2025).

A particularly significant finding concerns equity dimensions: low-achieving students experienced substantially larger treatment effects (d=0.92) than did high-achieving students (d=0.48). This pattern directly contradicts a common technology equity concern—that digital innovations often exacerbate existing disparities by providing the greatest benefits to already-advantaged learners. The opposite occurred in this study: AI-driven personalization appeared to preferentially benefit struggling learners, narrowing the achievement disparities.

This disparity-reducing pattern likely emerges because high-achieving students benefit less from additional personalization (having likely already received reasonably well-matched instruction in classroom contexts), while low-achieving students benefit substantially from targeted remediation and individualized pacing that traditional classroom instruction rarely provides. This effect is theoretically consonant with Vygotsky's ZPD concept: ZPD positioning becomes increasingly crucial for learners from existing competencies, while learners close to mastery benefit less from additional scaffolding (Setyoningrum et al., 2025).

The finding that effects did not substantially differ by gender is similarly encouraging for equity, suggesting that the technology did not reproduce or exacerbate gender disparities in academic achievement—an important consideration given the persistent gender gaps in mathematics and technical subjects across many educational contexts.

The somewhat attenuated effects in rural contexts (d=0.61) compared with urban contexts (d=0.84) warrant attention. This differential likely reflects infrastructure limitations (potentially less consistent Internet connectivity, less teacher experience with technology, and less robust technical support), rather than the fundamental unsuitability of AI-driven learning for rural students. The continued significance of the effects

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

across rural contexts suggests that with adequate infrastructural support and teacher professional development, rural students need not be excluded from AI-driven personalization benefits. However, implementation barriers in less-resourced contexts merit explicit policy attention.

An important theoretical and practical consideration concerns the potential trade-off between achievement and engagement. Concerning AI-driven learning systems, efficient pacing could theoretically produce achievement gains through accelerated instruction, at the cost of reduced motivation or engagement. The results do not support this trade-off hypothesis; instead, engagement increased alongside achievement in the experimental groups. This suggests that the system achieved what might be termed "optimal engagement"— a sufficient productive struggle to maintain psychological engagement while preventing boredom or frustration.

However, long-term engagement sustainability requires consideration beyond a 16-week intervention period. Initial engagement improvements may reflect novelty effects, with engagement declining following extended exposure. Longitudinal tracking beyond the intervention period would clarify whether engagement benefits persist and represent a valuable direction for future research.

The substantial experimental group advantage on transfer assessment items (d=0.66) was theoretically significant. Transfer performance demonstrates that improved achievement did not represent narrow task-specific learning, but rather a genuine conceptual understanding, enabling flexible application to novel contexts. This finding suggests that AI-driven personalization did not merely increase superficial memorization or procedural fluency but also promoted deeper conceptual understanding—arguably the most educationally valuable learning outcome.

The mechanism supporting transfer likely involves the system's capacity to systematically expose learners to problem and concept variations. Cognitive science research emphasizes that conceptual transfer requires exposure to multiple problem contexts, solution approaches, and concept instantiations. Traditional classroom instruction, constrained by limited class time and teaching resources, often presents limited problems and contextual variations. Conversely the AI system can generate nearly infinite problem variations while maintaining conceptual equivalence across surface features, systematically exposing learners to the problem space breadth necessary for transfer (Saberi et al., 2026).

The accelerating learning trajectories observed in the experimental groups warrant further discussion. The mixed-effects models revealed not only larger endpoint gains but also steeper learning slopes, suggesting that AI-driven benefits accumulated over time rather than representing initial advantages that subsequently plateaued. This trajectory pattern has important implications: continued AI-driven personalization may generate even larger cumulative effects over extended periods. Conversely, this pattern raises questions about trajectory sustainability: Would learning slopes continue accelerating indefinitely, or would trajectories eventually plateau at increasingly optimal mastery levels? A longer intervention duration would clarify this trajectory.

The initial equivalence between the experimental and control groups during weeks 1-4 followed by divergence thereafter suggests a "system adaptation" phase during which students developed proficiency with the AI platform before receiving full personalization benefits. This temporal pattern has practical implications. Educators implementing AI systems should anticipate an initial adaptation period during which achievement benefits may not be immediately apparent, requiring institutional patience and commitment through the adaptation phase.

The differential correlation between help-seeking and achievement across conditions (r=0.54 experimental, r=0.08) revealed important information about learning mechanisms. In the control conditions, students seeking help did not show corresponding achievement advantages, suggesting that available help resources provided limited instructional benefits, —possibly because help was generic rather than pedagogically targeted. Conversely, in the experimental conditions, help-seeking was positively correlated with achievement, suggesting that the AI system's pedagogically-sequenced hints genuinely facilitated learning. This finding implies that the value of personalized systems extends beyond content sequencing, to encompass personalized feedback and support structures.

Teachers' professional development. The 16-hour professional development provided to experimental teachers likely proved necessary for effective AI system utilization. Research on educational technology implementation consistently demonstrates that insufficient teacher preparation undermines intervention efficacy. However, 16 h may represent a relatively modest professional development investment compared to the magnitude of the effects achieved. Expanded professional development could potentially further enhance effects through improved learning analytics interpretation and classroom integration.

Technical Infrastructure and Support. Implementation across schools with varying technical infrastructure revealed that the effects persist even in less-resourced rural settings, yet somewhat attenuated effects in rural

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

contexts suggest that infrastructure investment facilitates more optimal outcomes. Cloud-based AI systems reduce local infrastructure demands compared with locally-hosted systems, potentially improving rural implementation feasibility.

Cost effectiveness Analysis. While this study did not conduct a formal cost-effectiveness analysis, the magnitude of achievement gains achieved through 180 minutes weekly of AI-driven learning (approximately 11.8% achievement improvement) appears economically reasonable compared with many educational interventions requiring substantially greater resource investment.

IV. CONCLUSIONS

This quantitative study examined the efficacy of an AI-driven adaptive learning system integrated into blended classroom environments in 12 Indonesian secondary schools. Through a rigorous experimental design involving 480 students assessed across multiple outcome dimensions, the investigation found that AI-driven adaptive learning substantially enhanced academic achievement (d=0.77), student engagement (d=0.64), adaptive skill development and transfer (d=0.66), and learning efficiency (23.5% time reduction). Importantly, the effects were equally distributed across diverse student populations, with low-achieving students experiencing the largest benefits, suggesting that AI-driven personalization narrows rather than exacerbates achievement disparities. Learning trajectory analysis revealed that benefits accumulated progressively over the 16-week intervention, with diverging growth slopes suggesting a sustained advantage for AI-supported learners. Subgroup analyses confirmed effect persistence across gender, initial achievement levels, and geographic contexts, although somewhat attenuated in rural settings, likely due to infrastructural limitations rather than fundamental unsuitability. These findings substantiate AI-driven adaptive learning as a promising educational innovation, with robust empirical support. However, the findings also emphasize that implementation quality, teacher professional development, and technical infrastructure substantially influence outcomes; AI systems represent enablers of improved learning, rather than panaceas. Future research addressing long-term sustainability, diverse AI system designs, and cost-effectiveness will continue to illuminate the role of AI-driven personalization in addressing these educational challenges. Specifically, for Indonesian education, this research suggests that AI-driven adaptive learning merits continued investment and policy support as a strategy for improving educational outcomes, particularly for underserved student populations.

Funding Statement

"No external funding was received for this study."

Ethical Compliance

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Data Access Statement

A Data Access Statement is a section in a scientific publication or research report that explains how the data used or generated in a study can be accessed by readers or other researchers. This statement aims to promote transparency, support research reproducibility, and comply with open-access policies, where applicable.

Common Elements in a Data Access Statement:

1. Data Location: Specifies where the data are stored, such as in online repositories (e.g., Zenodo, Dryad, or institutional repositories).
2. Access Instructions: Provides information on how to access the data, such as direct links, digital object identifiers (DOI), or contact details.
3. Data Availability: Indicates whether the data are publicly accessible, available upon request, or restricted due to ethical, legal, or privacy considerations.
4. Data Licensing: If the data are open, specify the applicable license (e.g., Creative Commons).

Examples of Data Access Statements:

1. Open Data:
    - "The data supporting this study are openly available in Zenodo at [DOI: 10.xxxx/zenodo.xxxx]."
2. Restricted Data:
    - "The data that support the findings of this study are available upon request from the corresponding author. Due to privacy concerns, the data are not publicly available."

Vol. 5 Issue 1, January-June 2026
e-ISSN: 2830-7933
DOI: 10.55299/ijere.v5i1.1807

3. No Data Available:
   - "No datasets were generated or analyzed during the current study."
4. Conditional Access:
   - "The data supporting this study are available under restricted access and can be obtained upon reasonable request to the corresponding author and with the permission of the ethics committee."

Purpose of a Data Access Statement:
- Reproducibility: Enables other researchers to replicate or verify the findings.
- Collaboration: Encourages further collaboration by sharing data.
- Compliance: Adheres to the policies of funding agencies or journals that require open access to data.

Conflict of Interest Declaration
The authors declare that they have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

ACKNOWLEDGEMENTS

The author thanks all people and institutions in most cases and the sponsor and financial support acknowledgments.

REFERENCES

Akour, I. A., Al-Maroof, R. S., Alfaisal, R., & Salloum, S. A. (2022). A conceptual framework for determining metaverse adoption in higher institutions of gulf areas: An empirical study using the hybrid SEM-ANN approach. *Computers and Education: Artificial Intelligence*, *3*: 100052. https://doi.org/10.1016/j.caeai.2022.100052

Ali, N., & Thue, D. (2025). Struggle Signals: Detecting Player Difficulties using Machine Learning. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, *21*(1), 175–185. https://doi.org/10.1609/aiide.v21i1.36821

Arikunto, S. (2016). *Metodologi Penelitian Suatu Pendekatan Proposal. Jakarta: Rineka Cipta.*

Atimoe, T. I., Iornem, K., Akogwu, R., Senkoya, M. (2025). rt The performance gap in motivational communication: An experimental test of linguistic framing onengagement and task accuracy. *International Journal of Science and Research Archive*, *17*(3), 797–801. https://doi.org/10.30574/ijsra.2025.17.3.3246

Balkist, P. S., Hamdallah, F., Kuncoro, K. S., Siswanto, R. D., Rahmat, D. (2025). Identifying the learning pace in mathematics classrooms: What teachers can see—and what they miss without information support? *Union: Jurnal Ilmiah Pendidikan Matematika* *13*(3), 949–960. https://doi.org/10.30738/union.v13i3.21915

Creswell, J. W. (2021). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.*

Dmytruk, A., Hrytsiv, V., Babkina, M., Skril, I., Vyslobodska, I., & Smilevska, M. (2025). Integration of artificial intelligence technologies into the digital transformation of professional higher education in technical fields. In *PROFESSIONAL EDUCATION AND PERSONNEL TRAINING* (pp. 110–143). TECHNOLOGY CENTER PC. https://doi.org/10.15587/978-617-8360-16-0.ch5

Khusnadin, M. H., Hariyanti, M., & Kadir, A. (2025). Ai-Driven Personalized Learning And Its Ethical Implications For Educational Counseling. *International Journal of Research in Counseling*, *3*(2), 150–164. https://doi.org/10.70363/ijrc.v3i2.263

Miles, M. B., & Huberman, A. M. (2014). *Qualitative data analysis: A methods sourcebook (3rd ed.). SAGE Publications*.

Montoya Espinoza, L. M., Coloma Chong, J. K., Nieto Herrera, D. J., & Quintuña Barrera, L. V. (2025). *Digital Transformation in Higher Education English Teaching: Technology Integration and Virtual Learning Environments*. RICAC Editorial 2025. https://doi.org/10.65488/9789942519238

Saberi, M., Adib, M.-E., Adib-Hajbaghery, M., Heidari, M. M., & Fallah, A. (2026). Developing and psychometrics assessment of a checklist for safe intrahospital patient transfer. *BMC Health Services Research*. https://doi.org/10.1186/s12913-025-13987-w

Setyoningrum, T. Y., Susanto, D. A., & Setiaji, A. (2025). IMPACTING ARTIFICIAL INTELLIGENT CHATBOT IN FLIPPED CLASSROOM TO ENHANCE STUDENTS' SPEAKING COMPETENCE. *Applied Research on English Education (AREE)*, *3*(1), 8. https://doi.org/10.26714/aree.3.1.2025.8-25

Sugiyono. (2019). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R&D. Alfabeta.*

Thirusanku, D. J., & Raman, K. R. T. (2025). AI-Driven Marketing Personalization. *Account and Financial Management Journal*, *10*(11), 3843–3850. https://doi.org/10.47191/afmj/v10i11.04