Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluating AI-Driven Adaptive Gamification Using a Documented Synthetic Educational Dataset (SynEduGam-10K): A Reproducible Analytical Framework

Pauly Awad
American University of the Middle East
Soraia Oueida
American University of the Middle East
Claudia Mattar
claudia.mattar@balamand.edu.lb

University of Balamand

Research Article

Keywords: Artificial Intelligence, Education, Synthetic Data, Adaptive Gamification, Learning Analytics, Machine Learning

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11754439/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The use of Artificial Intelligence (AI) in education has made it possible to create learning environments that adapt and change based on data. However, getting real educational data is challenging because of privacy rules, ethical issues, and problems between organizations. This study presents SynEduGam-10K, a synthetic dataset with 10,000 learners, including information about their actions, how motivated they are, their school performance, and how they interact with game-like features. The dataset uses equations to generate the data. It also shows a way to test AI-based games. Random Forest and XGBoost models were used to predict grades using only the actions and game features. These models had an R^2 score of up to 0.74 (RMSE = 4.78). They could also classify performance into three groups with up to 69.5% accuracy. The errors happened mostly when the groups were close to each other. A multiple regression showed that 76% of the results could be explained with engagement ($\beta = 0.64$) and motivation ($\beta = 0.58$) as factors. The analysis showed that study time had no effect on the results unless it was connected to engagement. A transparent simulation that increased engagement levels across the data showed a significant improvement in results when using adaptive gamification instead of regular gamification. The change was 1.20 points ($t(19,984) = 9.01$; $p < .001$; Cohen's $d = 0.13$). This shows that testing ideas before using them in real-world can save money and time. It also shows that focusing on engagement might not be enough if motivation is also important for success. The dataset, the generative equations used to create it, and the full R code for analysis are released to support benchmarking and replication. This helps researchers test their ideas, repeat the work, and do experiments without risking people's privacy in AI-based education.

1. Introduction

Artificial Intelligence (AI) has emerged as a revolutionary influence in modern education, enabling customized learning routes, predictive analytics, and smart tutoring systems [1, 2]. With the rapid development of digital learning platforms, the amount of learner interaction data for analysis has increased proportionately, and comprehensive studies illustrate a swiftly growing body of evidence about AI applications in educational fields [3, 4, 5]. At the same time, the field encounters an ongoing structural limitation: acquiring and disseminating high-quality educational datasets is challenging due to privacy regulations like the General Data Protection Regulation (GDPR), institutional data governance policies, and valid ethical apprehensions regarding the secondary utilization of student data [6, 5]. These limitations restrict methodological advancement, obstruct replication efforts, and centralize research capabilities among institutions that possess extensive proprietary datasets.

The development of synthetic data has surfaced as a viable solution to this limitation. Synthetic datasets facilitate open experimentation, benchmarking, and education by generating fictitious records that maintain the statistical integrity of a target domain without revealing any actual individuals, hence adhering to privacy regulations [7, 8, 9]. Research in related fields indicates that models developed using carefully designed synthetic data can achieve predictive performance comparable to those trained on actual data [10], and recent investigations have started to assess synthetic tabular data generation

specifically in the context of learning analytics [11]. Nonetheless, researchers still have limited access to publicly available synthetic datasets that encompass behavioral, motivational, and gamification factors for educational research.

Simultaneously, gamification, which is the incorporation of game design components in non-gaming environments [12], has been extensively utilized to enhance learner motivation and involvement, with meta-analytic findings demonstrating favorable yet varied impacts on educational results [13] and recent empirical research validating advantages for performance and engagement in real-world courses [14, 15]. A notable deficiency of first-generation implementations is their rigidity: all learners are awarded identical points, badges, and leaderboards irrespective of personal requirements, which constrains sustained efficacy [16]. Another research on customized or adaptive gamification advocates for the personalization of game components according to learner traits, although it acknowledges that the empirical foundation for such customization remains limited and methodologically inconsistent [17].

Assessing adaptive gamification thoroughly presents significant costs: it necessitates equipped platforms, ongoing participant groups, and ethical approval for conducting experiments with learners. A recorded synthetic dataset provides an alternative approach to a more economical, privacy-conscious setting where analytical frameworks, forecasting models, and intervention strategies can be created, refined, and rigorously tested prior to real-world application. This research follows that trajectory. It presents SynEduGam-10K, a synthetic dataset of 10,000 simulated learners with variables interconnected by well-defined generative equations, and uses it to illustrate a comprehensive, fully repeatable assessment workflow executed using statistical software R.

The analysis is directed by four research questions (RQs):

- • **RQ1.** What are the statistical and structural characteristics of SynEduGam-10K, and to what extent do they accurately represent their recorded generative design?
- • **RQ2.** To what extent can ensemble machine learning models (Random Forest, XGBoost) precisely forecast simulated learner success based solely on behavioral and gamification attributes?
- • **RQ3.** What learner attributes are the most significant predictors of simulated academic achievement?
- • **RQ4.** Given transparent and clearly articulated assumptions, what impact does a simulated adaptive gamification intervention have on results compared to a fixed baseline?

In contrast to research that presents actual results, this study regards the adaptive-versus-static comparison solely as a simulated scenario, with all assumptions well-articulated and parameters adjustable in the provided code. The contribution is thus both methodological and infrastructural: a documented dataset, a clear generative specification, and a comprehensive open analysis pipeline that can be replicated, inspected, and enhanced by other researchers.

2. Literature review

2.1 Artificial Intelligence in Education

AI applications in education (AIEd) have progressed many paradigms, ranging from AI-driven systems that provide predetermined content to AI-facilitated collaboration and ultimately to AI-enhanced frameworks where learners take control over intelligent tools [2]. Holmes and Tuomi [1] present a classification of AIEd systems that includes applications for students, teachers, and institutions, while warning that aspirations often exceed technological capabilities. Systematic evaluations reveal a shared array of prospects, personalized education, automated evaluation, and predictive analytics, coupled with persistent issues of ethics, educator readiness, and insufficient theoretical foundations [4, 5]. The latest meta-analysis of higher education literature highlights profiling and prediction, intelligent tutoring, assessment, and adaptive systems as the primary application categories while urging enhanced rigor, collaboration, and ethical considerations in AIEd research [3].

In the realm of predictive analytics, machine learning models designed to forecast student performance are among the most extensively researched applications of AIEd. [18] Systematic reviews indicate that tree-based ensembles, especially Random Forest, alongside support vector machines and neural networks, are the most utilized and highest-performing algorithmic families for this purpose, utilizing demographic, behavioral, and assessment characteristics [19, 20]. The emergence of extensive language models has introduced enhanced functionalities for content creation and assessment, although it has also brought out novel practical and ethical dilemmas, particularly with the privacy of learner data utilized in model construction and evaluation [6]. This conflict over privacy serves as a primary impetus for the current research.

2.2 Gamification and Adaptive Gamification in Learning

Gamification is traditionally characterized as the application of game design components such as points, badges, levels, leaderboards, and challenges in settings outside of gaming [12]. Initial empirical assessments revealed predominantly favorable, although contextually contingent impacts on engagement and motivation [16]. Meta-analytic findings have reinforced this perspective: gamification has notable small-to-moderate impacts on cognitive, motivational, and behavioral learning results, exhibiting considerable variability across different designs and contexts [13]. Recent educational research consistently demonstrates quantifiable advantages; for instance, the integration of digital and non-digital gamification has enhanced learning outcomes, cognitive involvement, and course satisfaction in language education [15], while a three-year longitudinal study revealed that gamified learning correlated with superior academic achievement compared to both online and conventional methods [14].

The variability of gamification outcomes has spurred investigations into customized gamification, where game components are modified according to learner characteristics, inclinations, or actions. A comprehensive analysis of the literature indicates that most customization initiatives depend on fixed player classifications, that experimental data, especially regarding learning outcomes, is often insufficiently robust, and that the discipline requires more rigorous methodological frameworks and

collaborative assessment tools [17]. AI-powered adaptive gamification evolves from singular customization to ongoing, data-informed modifications of challenge levels, incentives, and responses according to anticipated learner conditions. Due to the complexity of constructing and assessing such systems in vivo, reusable datasets and clear evaluation frameworks serve as essential community resources, this study aims to bridge that gap.

2.3 Synthetic Data in Educational Research

Synthetic data, defined as artificially created datasets intended to emulate the statistical characteristics of actual data, has evolved from a method for statistical disclosure control into a widely adopted instrument for privacy-preserving machine learning [7, 8, 9]. Generation methodologies include explicit probabilistic frameworks, as demonstrated in this context, as well as deep generative architectures like GANs, variational autoencoders, diffusion models, and, most recently, extensive language models for tabular generation [21]. Evidence from privacy-sensitive sectors indicates that carefully constructed synthetic data can facilitate model advancement with minimal utility degradation [10].

In the realm of education, Liu et al. [11] assessed cutting-edge tabular generators applied to learning analytics data and discovered that synthetic datasets maintain both similarity and subsequent predictive efficacy while minimizing disclosure risk, thus proving synthetic data to be an effective tool for expanding learning analytics research. The explicit, equation-driven generation utilized for SynEduGam-10K sacrifices a degree of distributional authenticity for total transparency: all dependencies within the data are identifiable, examinable, and modifiable, rendering these datasets especially appropriate for benchmarking analytical processes, educational purposes, and simulation research where the ground truth is essential. A comprehensive educational dataset that encompasses behavioral, motivational, academic, and gamification characteristics, along with open-source code for a complete evaluation framework for adaptive gamification research, has been lacking.

2.4 Research Gap

Integrating the previously listed academic pillars: (a) AIEd research necessitates a data architecture that safeguards privacy [3, 6]; (b) Research on adaptive gamification is deficient in common assessment resources and lacks methodologically sound performance evidence [17]; and (c) while synthetic data methodologies are prepared for educational use, documented cases in education are still limited [11]. This research tackles the convergence of these deficiencies by providing a documented synthetic dataset with an established generative framework and illustrating, comprehensively, how it facilitates statistical validation, predictive modeling, and transparent simulation of an adaptive gamification intervention.

3. Research Contribution

This research presents four significant contributions:

- • A comprehensive synthetic educational dataset (SynEduGam-10K; N = 10,000) with all generating equations, parameters, and derived variables transparently provided;
- • A replicable R workflow that reconstructs each table, statistic, and figure presented herein from the original data utilizing a constant random seed;
- • A truthful empirical assessment of ensemble model efficacy in forecasting simulated outcomes based solely on behavioral and gamification attributes, incorporating a leakage-conscious feature design;
- • A clear Monte Carlo approach for evaluating adaptive gamification situations under stress, where intervention assumptions are defined and modifiable parameters instead of unspoken assertions.

4. Methodology

4.1 Dataset Structure and Generative Design

SynEduGam-10K consists of 10,000 simulated student profiles featuring ten variables that include behavior (study hours), psychological state (motivation), platform interaction (engagement), assessment (quiz and assignment scores), outcomes (final grade), and gamification telemetry (points, badges, level). The relationships among variables are established via the generative equations in Table 1: weekly study hours adhere to a truncated normal distribution; engagement is a stochastic linear function of study hours; motivation is produced independently; assessment scores combine engagement and motivation with additive noise; and the final grade is a weighted aggregation of quiz score, assignment score, and engagement. Gamification elements are extracted from user activity to replicate platform telemetry. All scoring variables are represented on arbitrary continuous scales instead of percentage-of-maximum metrics; linear rescaling to institutional grading standards does not alter any of the standardized statistics presented below.

Table 1
Variables and generative definitions in SynEduGam-10K.

Variable	Generative definition	Observed range
Study hours	~ Normal (μ = 20, σ = 8), truncated to [5, 40]	5.0–40.0
Engagement	0.5 × study hours + ε, ε ~ Normal (0, 10), floored at 0	0.0–47.3
Motivation	~ Normal (μ = 60, σ = 15), truncated	5.2–100.0
Quiz score	0.6 × engagement + 0.4 × motivation + ε (σ ≈ 8.0)	0.0–76.5
Assignment score	0.5 × engagement + 0.5 × motivation + ε (σ ≈ 8.0)	0.0–78.2
Final grade	0.4 × quiz + 0.4 × assignment + 0.2 × engagement	0.0–65.3
Points	Derived from engagement (platform telemetry)	0.0–523.2
Badges	Near-constant in this release (2/10,000 non-zero)	0–1
Level	Derived from cumulative points	0.0–6.3
Performance class	As distributed: 98.7% Low, 1.3% Medium, 0% High	categorical

Note. Residual noise standard deviations for quiz and assignment scores (≈ 8.0) were empirically recovered from the distributed data and matched the generative design. ε terms are independent draws.

4.2 Data Screening and Derived Variables

Initially, the categorical performance class variable exhibits significant imbalance (98.7% “Low,” 1.3% “Medium,” and 0 instances of “High”), hence preventing substantial group comparison or categorization. Consequently, for all analyses conducted on groups, a balanced three-tier performance metric derived from final-grade tertiles (cut points 24.39 and 32.49; n = 3,334 / 3,333 / 3,333 for Low / Medium / High respectively) was established. Secondly, the badges variable is nearly invariant (9,998 out of 10,000 values are zero) and therefore excluded from the modeling. Both resolutions are executed and annotated in the published analytical script.

4.3 Statistical Validation Strategy

Dataset validation tackled RQ1 at three different levels: (a) distributional fidelity, which involved comparing the observed moments (mean, dispersion, skewness) of each variable with its generative specification; (b) structural fidelity, which ensured that the observed correlation matrix accurately reflects the dependency structure suggested by the generative equations, encompassing the intended engagement–outcome and motivation–outcome pathways; and (c) utility validation, which verified that supervised models can retrieve the established predictive relationships. All analyses were conducted using R version 4.6.0 with a constant global random seed (123).

4.4 Machine Learning Models

Two ensemble learning methods were assessed: Random Forest [22] (100 trees, default mtry) and XGBoost [23] (learning rate 0.1, maximum depth of 6, 150 boosting iterations). The dataset was divided into training and testing subsets in an 80/20 ratio. To prevent target leakage, scores from quizzes and assignments were omitted from the feature set, as the final grade is derived from them; thus, models solely forecast outcomes based on behavioral and gamification attributes (study hours, engagement, motivation, points, and level). Two assignments were evaluated on the reserved test dataset: (a) regression concerning the continuous final grade, represented as RMSE, MAE, and R^2; and (b) classification of the resultant performance tertile, shown by accuracy and macro-averaged F1, with confusion matrices analyzed for error patterns. Permutation importance (%IncMSE) was utilized to derive feature significance for Random Forest, while gain was employed for XGBoost.

4.5 Simulated Adaptive Gamification Scenario

RQ4 was examined using a Monte Carlo simulation, with all assumptions fully articulated. The static state refers to the documented final-grade distribution of SynEduGam-10K. The adaptive condition is a hypothetical scenario developed in two phases: (a) a categorized relative engagement increase is implemented, focusing adaptive assistance on lower-performing students (+ 35%, + 20%, and + 10% engagement for the Low, Medium, and High performance tertiles, respectively), embodying the principles of adaptive systems that distribute support, reward pacing, and feedback intensity based on anticipated needs; and (b) the enhanced engagement is transmitted through the dataset's generative equations to yield contrasting values for quizzes, assignments, and final grades, utilizing residual noise standard deviations derived from the data. The evaluation of the conditions employed Welch's t-test, Cohen's d (along with 95% confidence intervals), and the mean difference together with its confidence interval. The uplift percentages are hypothetical scenarios and modifiable variables in the published script, not empirical assessments; the simulation illustrates the evaluation process and quantifies the consequences of the specified assumptions and is understood solely in this context.

5. Results

5.1 Descriptive Statistics and Distributional Validation (RQ1)

Table 2 presents descriptive statistics for the six fundamental continuous variables. The recorded instances align closely with the generative specification: average study hours are 20.06 (SD = 7.76) compared to a defined Normal (20, 8); motivation averages 59.80 (SD = 14.83) against Normal (60, 15); and engagement averages 11.13, in accordance with its characterization as half of study hours plus zero-mean noise. The skewness is moderate throughout ($|\text{skew}| \leq 0.62$), and the final grade exhibits near symmetry (skew = 0.18). Figure 1 illustrates the distribution of final grades: unimodal, approximately normal, with a minor right skew, an internally coherent manifestation of the composite outcome equation.

Table 2
Descriptive statistics for the core continuous variables (N = 10 000).

Variable	Mean	SD	Median	Min	Max	Skew
Study hours	20.06	7.76	19.98	5.00	40.00	0.12
Engagement	11.13	9.19	10.08	0.00	47.29	0.62
Motivation	59.80	14.83	59.91	5.17	100.00	−0.01
Quiz score	30.55	11.48	30.31	0.00	76.46	0.09
Assignment score	35.52	11.93	35.39	0.00	78.17	0.04
Final grade	28.66	9.26	28.34	0.00	65.31	0.18

Note. All values were computed from the distributed dataset using the released R script. Scores are on arbitrary continuous scales.

5.2 Correlation Structure (RQ1)

Table 3 and Fig. 2 illustrate the Pearson correlation matrix. The examined structure accurately reflects the generative design. Assessment scores exhibit the highest correlation with the final grade (quiz r = .87; assignment r = .86), as anticipated based on its numerical components. Among the upstream factors, engagement (r = .65) and motivation (r = .60) are the principal correlates of success, although study hours have a moderate association with the final grade (r = .23) and mainly influence it through engagement (study hours–engagement r = .35). Motivation is inherently unrelated to study behavior (r = .03), providing the dataset with two distinct routes to achievement, behavioral and motivational, that subsequent models must differentiate. Figure 3 illustrates the affirmative, although irregular, study-time gradient (r = .23, p < .001).

Table 3
Pearson correlations among core variables (N = 10 000; all |r| ≥ .03 are significant at p < .01).

Variable	1	2	3	4	5	6
1. Study hours	—					
2. Engagement	.35	—				
3. Motivation	.03	.03	—			
4. Quiz score	.17	.50	.53	—		
5. Assignment score	.14	.40	.64	.53	—	
6. Final grade	.23	.65	.60	.87	.86	—

5.3 Group Differences in Engagement (RQ1, RQ3)

Since the performance categories are established based on the final grade, examining grade disparities among classifications would be redundant; the essential inquiry is whether the performance groups exhibit variations in engagement behavior. A one-way ANOVA revealed a substantial, statistically significant impact of performance group on involvement, $F(2, 9997) = 2{,}707$, $p < .001$, $\eta^2 = .35$ (Table 4). Tukey HSD analyzes and validates that all pairwise distinctions are statistically significant ($p < .001$): compared to the Low group, the average engagement is 4.93 points greater in the Medium group and 13.20 points greater in the High group. Figure 4 illustrates the resultant differentiation in final-grade distributions among various categories.

Table 4
One-way ANOVA: engagement across derived performance tertiles.

Source	df	SS	MS	F	p
Performance group	2	296,358	148,179	2,707	< .001
Residual	9,997	547,306	55		

Note. $\eta^2 = .35$. Tukey HSD: Medium − Low = 4.93 [4.51, 5.36]; High − Low = 13.20 [12.77, 13.62]; High − Medium = 8.26 [7.84, 8.69]; all $p < .001$.

5.4 Predictors of Simulated Achievement (RQ3)

A multiple linear regression analyzing final grades based on study hours, engagement, and motivation explained approximately 76% of the variance in outcomes, $F(3, 9996) = 10{,}520$, $p < .001$, $R^2 = .759$ (Table 5). Engagement ($B = 0.641$, $\beta = .64$) and motivation ($B = 0.361$, $\beta = .58$) emerged as strong independent predictors. Significantly, study hours had no substantial direct impact after accounting for involvement ($B = -0.009$, $p = .133$, $\beta \approx -.01$): in line with the generative framework, the effect of study duration on performance is entirely mediated by engagement. This pattern serves as a significant benchmark property; analysis pipelines utilized on this dataset are expected to produce complete mediation, and any failure to do this indicates a specification issue.

Table 5
Multiple regression predicting final grade (N = 10 000)

Predictor	B	SE	β	t	P
Intercept	0.121	0.220	—	0.55	.583
Study hours	−0.009	0.006	−.01	−1.50	.133
Engagement	0.641	0.005	.64	121.46	< .001
Motivation	0.361	0.003	.58	117.80	< .001

Note. $R^2 = .759$, adjusted $R^2 = .759$; $F(3, 9996) = 10{,}520$, $p < .001$. β = standardized coefficients. 95% CIs: engagement [0.630, 0.651]; motivation [0.355, 0.367].

5.5 Predictive Modeling Performance (RQ2)

Table 6 represents the performance of both models and tasks on the held-out test set; Fig. 5 illustrates the significance of features. Predicting the continuous final grade only based on behavioral and gamification attributes, XGBoost attained RMSE = 4.78, MAE = 3.82, and R^2 = .739, slightly surpassing Random Forest (RMSE = 4.85, MAE = 3.87, R^2 = .731); both models are close to R^2 = .759. The ceiling suggested by the noise configuration of the generative equations signifies that the ensembles retrieve almost all attainable signals. In the three-class tertile challenge, XGBoost outperformed Random Forest with an accuracy of 69.5% and a macro-F1 score of .696, compared to Random Forest's 68.6% and .687, significantly exceeding the 33.3% chance level for balanced classes. The confusion matrix indicates that inaccuracies predominantly occur between neighboring tertiles, with merely 22 out of 2,000 test instances (1.1%) misclassified between the Low and High extremes. Both models recognized engagement as the primary predictor, followed by gamification telemetry derived from engagement (points, level) and motivation, reflecting the regression analysis results.

Table 6
Held-out test performance (20% split) for both models and tasks

Model	RMSE	MAE	R^2	Accuracy	Macro-F1
Random Forest	4.85	3.87	.731	.686	.687
XGBoost	4.78	3.82	.739	.695	.696

Note. RMSE, MAE, R^2: regression on final grade. Accuracy, macro-F1: classification of derived performance tertile (chance = .333). The features include study hours, engagement, motivation, points, and levels.

5.6 Simulated Adaptive versus Static Gamification (RQ4)

According to the specified assumptions (Section 4.5), the tiered engagement enhancement increased average engagement by 17% and resulted in a mean simulated final-grade enhancement of 1.20 points (4%) compared to the static condition (Table 7; Fig. 6). The difference is statistically significant, Welch t (19,984) = 9.01, $p < .001$, 95% CI of the difference [0.94, 1.46]; yet it has little practical implication, with Cohen's d = 0.13, 95% CI [0.10, 0.16]. The mechanism is enlightening, although motivation plays a significant and independent role in achievement within this data-generating framework; even a big engagement-focused intervention results in slight changes in outcomes. The simulation thus measures a design insight for adaptive gamification; personalization tactics aimed at a singular behavioral channel may fall short when success is influenced by multiple factors, while illustrating how these situations can be analyzed economically and transparently prior to field tests.

Table 7
Simulated comparison of static and adaptive gamification conditions (N = 10 000 per condition)

Condition	Mean	SD
Static gamification (observed data)	28.66	9.26
Adaptive gamification (simulated)	29.85	9.51

Note. Δ = 1.20, 95% CI [0.94, 1.46]; Welch t (19,984) = 9.01, p < .001; Cohen's d = 0.13, 95% CI [0.10, 0.16]. The adaptive condition is a Monte Carlo counterfactual generated under the explicitly stated uplift assumptions of Section 4.5, not an empirical intervention effect.

6. Discussion

Three results deserve highlighting. Initially, the dataset operates precisely as described (RQ1): the observed moments, correlations, and the retrieved residual noise configuration all align with the generative framework. This fidelity is the characteristic that renders the dataset valuable as community infrastructure since the ground truth is established; SynEduGam-10K can function as a verification benchmark where analytical pipelines are anticipated to retrieve particular, pre-registered structural facts (for instance, the complete mediation of study time by engagement illustrated in Section 5.4). Evaluation criteria with established answers are commonplace in machine learning, although they are infrequently utilized in learning analytics, where they could mitigate the reproducibility issues highlighted in recent meta-analyses [3].

Secondly, the outcomes of the predictive modeling (RQ2, RQ3) provide a genuine reference for the literature on student performance prediction. When assessment scores are removed to prevent target leakage, ensemble models forecast performance based on behavioral and gamification attributes with $R^2 \approx .73–.74$ and categorize performance tertiles at approximately 69% accuracy, which is strong yet significantly lower than the near-ceiling accuracies occasionally documented in studies that incorporate elements of the target itself [19, 20]. The use of engagement and its related telemetry to determine feature significance aligns with the widely accepted view in learning analytics that interaction behavior is the most useful early indicator of success [24]. This suggests that consistently tracked gamification telemetry (such as points and levels) can stand in for engagement when direct measures are unavailable. Third, the simulation (RQ4) demonstrates the potential and the rigor of scenario analysis. The adaptive condition yielded a statistically significant yet modest overall effect (d = 0.13), with a substantial 35% increase in engagement for the lowest tertile. The underlying issue is structural: motivation plays a distinct role in achievement; hence, an intervention focused just on engagement neglects a significant factor. This mechanism aligns with the literature on tailored gamification, which increasingly posits that effective personalization should consider various learner attributes rather than focusing on a singular behavioral metric [17] and is supported by meta-analytic findings indicating that the impact of gamification on learning outcomes is generally modest and influenced by design decisions [13]. The exercise methodologically demonstrates how transparent simulation transforms ambiguous

anticipations regarding adaptive systems into measurable, clearly defined predictions that can be analyzed and modified prior to the enrollment of any student in a trial.

The research illustrates a comprehensive procedure for conducting privacy-preserving AIEd studies: articulate a generative design, disseminate the data and code, verify accuracy, and distinguish empirical assertions (characteristics of the dataset and models) from simulated assertions (outcomes of specified assumptions). This separation directly tackles the ethical issues surrounding learner data that drive synthetic methodologies [11, 6] while ensuring the epistemic clarity of each reported figure.

7. Implications

For academics and researchers, SynEduGam-10K and its framework offer a complimentary, privacy-conscious setting for the creation and validation of learning analytics techniques, evaluating student-performance models in leakage-aware scenarios, and designing adaptive-intervention assessments with established ground truth. For educational technology developers, the simulation findings warn against only engagement-driven personalization: adaptive gamification systems must integrate motivational assessments and multi-channel strategies to significantly enhance accomplishment. The discovery that gamification telemetry serves as a proxy for engagement indicates a cost-effective early-warning mechanism readily available in numerous gamified platforms for educators and institutions. The release model, which contains specified generative equations and executable analysis code, provides the community with a framework for reproducible, ethically sound AIEd research that aligns with recent demands for increased rigor [3].

8. Limitations

The results should be interpreted within defined boundaries. Primarily, SynEduGam-10K is synthetic: its linear-Gaussian generative framework excludes nonlinearities, feedback mechanisms, temporal variations, measurement inaccuracies, and demographic diversity present in genuine learning, thus rendering absolute performance metrics inapplicable to real-world contexts, and no significant educational assertions should rely solely on this dataset. Secondly, the comparison between adaptive and static approaches is a simulation that derives its implications from specified assumptions; it illustrates an assessment methodology and the inherent reasoning of the data-generating mechanism, rather than the practical effectiveness of adaptive gamification, a matter that can only be resolved through field experiments. Third, the gamification variables are analytically extracted from engagement instead of being recorded from authentic game mechanics, and one distributed variable (badges) proved unusable, while the distributed performance categories necessitated re-derivation from final-grade tertiles; both handling decisions, however documented, are at the discretion of the analyst. Fourth, engagement and motivation are depicted as singular continuous variables, which simplifies complex multidimensional entities. Ultimately, the outcomes depend on the hyperparameters and the singular 80/20 split illustrated, while the fixed-seed code provided ensures that every figure is precisely reproducible and facilitates the exploration of different options.

9. Conclusion

This research presented SynEduGam-10K, a comprehensive synthetic educational dataset, and showcased a completely repeatable R workflow for verifying the dataset, evaluating leakage-aware predictive models, and rigorously testing an adaptive gamification scenario based on certain assumptions. Ensemble models captured almost all recoverable signals (R^2 reaching .74 from behavioral and gamification attributes). Engagement was identified as the central element, mediating study duration and represented by gamification telemetry, and a clear simulation demonstrated that engagement-focused adaptive gamification produces only slight simulated improvements when motivation independently influences success. The study enhances reusable, privacy-conscious infrastructure for AI-focused educational research by disseminating data, generative equations, and comprehensive analysis code, while also providing a framework for differentiating between empirical and simulated assertions, an increasingly vital distinction as the discipline expands.

10. Future Work

Numerous extensions arise immediately from this research. (a) External validation: the methodology must be utilized on authentic, publicly licensed educational datasets, such as extensive learning-management-system corpora, to measure the disparity in utility between synthetic and actual data for both prediction and simulation. (b) Future iterations of SynEduGam ought to integrate nonlinear dependencies, longitudinal session-level data, demographic characteristics for equity assessments, and genuine gamification elements (badge regulations, streaks, leaderboard dynamics), while also contrasting the current equation-centric methodology with deep generative and LLM-driven tabular synthesis [21] under established privacy protections like differential privacy [7]. (c) Simulation depth: the intervention framework must progress from a singular engagement enhancement to a closed-loop simulation where a predictive model determines interventions for each period, allowing for the accumulation of impacts and facilitating in-silico evaluation of adaptive strategies prior to implementation. (d) Human validation: specialist educators might evaluate the credibility of simulated learner profiles, thus enhancing face validity with statistical accuracy. (e) Field translation: the hypotheses refined in this context that multi-channel personalization aimed at both engagement and motivation surpasses engagement-only adaptation will need to be examined in registered classroom experiments, thereby bridging the gap between synthetic prototyping and empirical validation [14, 17]. (f) Community benchmarking: we encourage replication and expansion; since the ground truth is established, SynEduGam-10K can serve as a reference for collaborative efforts where various analytical frameworks are evaluated based on their proficiency in retrieving described structures.

Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Funding

The authors did not receive support from any organization for the submitted work.

Author Contribution

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Pauly Awad and Soraia Oueida. All authors read and approved the final manuscript.

Data Availability

The datasets generated during and analyzed during the current study are available from the corresponding author upon request.

References

1. Holmes W, Tuomi I. State of the art and practice in AI in education. Eur J Educ. 2022;57(4):542–70.
2. Ouyang F, Jiao P. Artificial intelligence in education: The three paradigms. Computers Education: Artif Intell. 2021;2:100020.
3. Bond M, Khosravi H, De Laat M, Bergdahl N, Negrea V, Oxley E, Siemens G. A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. Int J educational Technol High Educ. 2024;21(1):4.

4. Chiu TK, Xia Q, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers Education: Artif Intell. 2023;4:100118.
5. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education–where are the educators? Int J educational Technol High Educ. 2019;16(1):39.
6. Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, Gašević D. Practical and ethical challenges of large language models in education: A systematic scoping review. Br J Edu Technol. 2024;55(1):90–112.
7. Drechsler J, Haensch AC. 30 years of synthetic data. Stat Sci. 2024;39(2):221–42.
8. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., …Weller, A. (2022). Synthetic Data–what, why and how? *arXiv preprint arXiv:2205.03257*.
9. Nikolenko SI. (2019). Synthetic data for deep learning.
10. Yale A, Dash S, Dutta R, Guyon I, Pavao A, Bennett KP. Generation and evaluation of privacy preserving synthetic health data. Neurocomputing. 2020;416:244–55.
11. Liu Q, Khalil M, Jovanovic J, Shakya R. (2024, March). Scaling while privacy preserving: A comprehensive synthetic tabular data generation and evaluation in learning analytics. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*(pp. 620–631).
12. Deterding S, Dixon D, Khaled R, Nacke L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*(pp. 9–15).
13. Sailer M, Homner L. The gamification of learning: A meta-analysis. Educational Psychol Rev. 2020;32(1):77–112.
14. Lampropoulos G, Sidiropoulos A. Impact of gamification on students' learning outcomes and academic performance: A longitudinal study comparing online, traditional, and gamified learning. Educ Sci. 2024;14(4):367.
15. Qiao S, Yeung SSS, Zainuddin Z, Ng DTK, Chu SKW. Examining the effects of mixed and non-digital gamification on students' learning performance, cognitive engagement and course satisfaction. Br J Edu Technol. 2023;54(1):394–413.
16. Hamari J, Koivisto J, Sarsa H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*(pp. 3025–3034). Ieee.
17. Oliveira W, Hamari J, Shi L, Toda AM, Rodrigues L, Palomino PT, Isotani S. Tailored gamification in education: A literature review and future agenda. Educ Inform Technol. 2023;28(1):373–406.
18. Baker RS, Martin T, Rossi LM. (2016). Educational data mining and learning analytics. The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications, 379–396.

19. Albreiki B, Zaki N, Alashwal H. A systematic literature review of student'performance prediction using machine learning techniques. Educ Sci. 2021;11(9):552.
20. Namoun A, Alshanqiti A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Appl Sci. 2020;11(1):237.
21. Shi R, Wang Y, Du M, Shen X, Chang Y, Wang X. (2025). A comprehensive survey of synthetic tabular data generation. *arXiv preprint arXiv:2504.16506*.
22. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
23. Chen T, Guestrin C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*(pp. 785–794).
24. Siemens G, Baker RSD. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*(pp. 252–254).

Figures

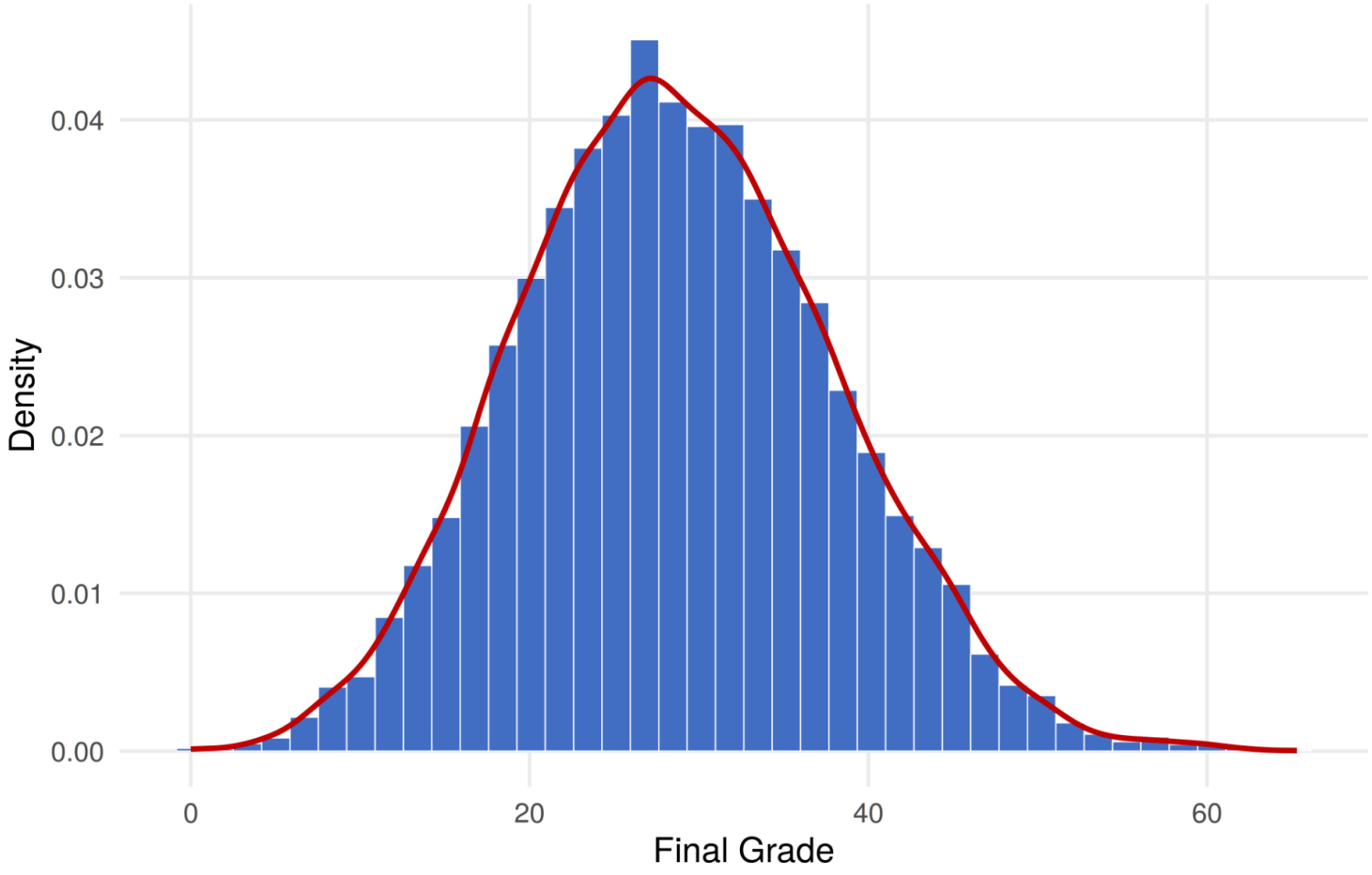


Figure 1

Distribution of final grades in SynEduGam-10K

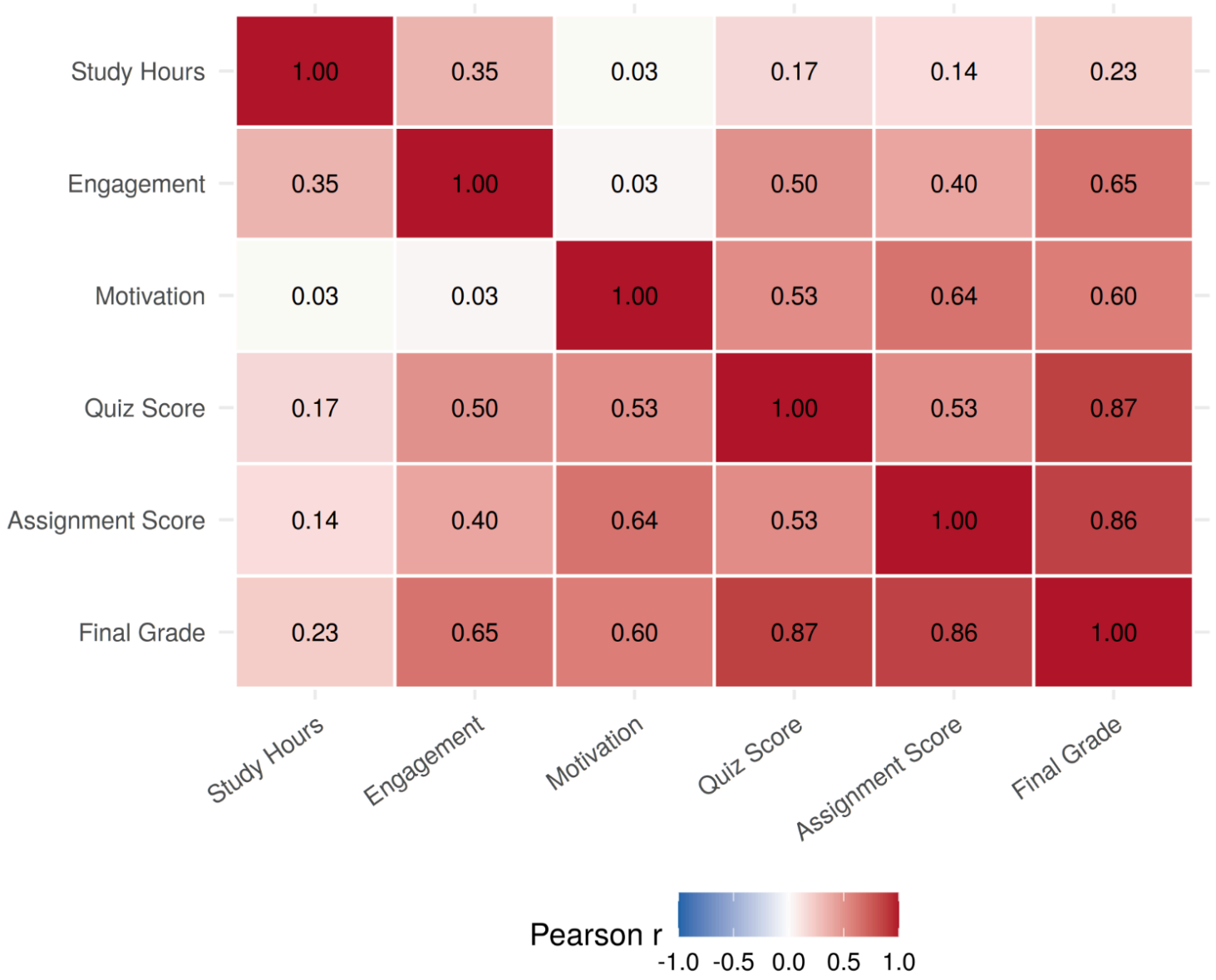


Figure 2

Correlation heatmap of the six core variables

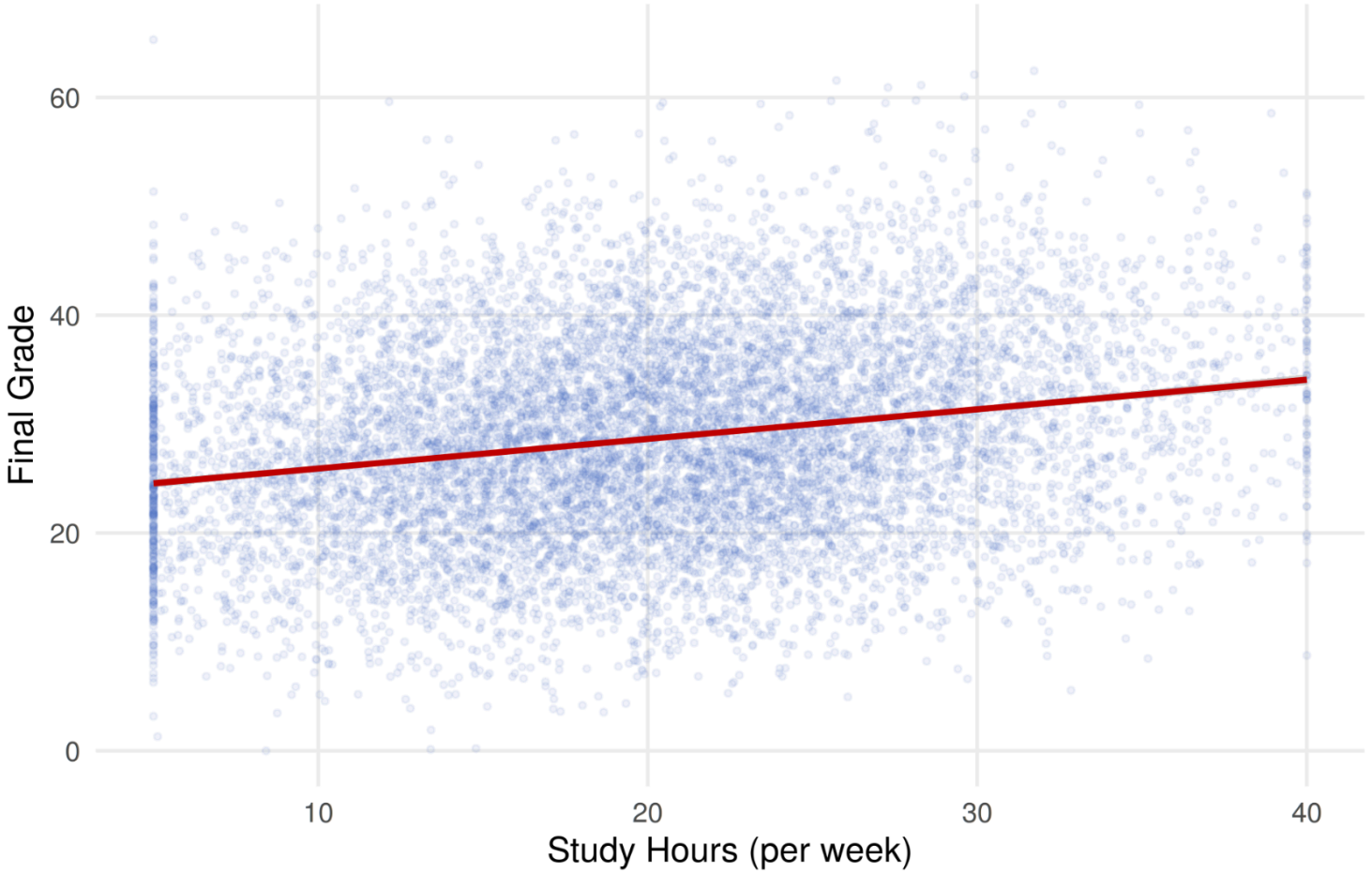


Figure 3

Relationship between weekly study hours and final grade (points: individual stimulated learners; line: OLS fit with 95% confidence)

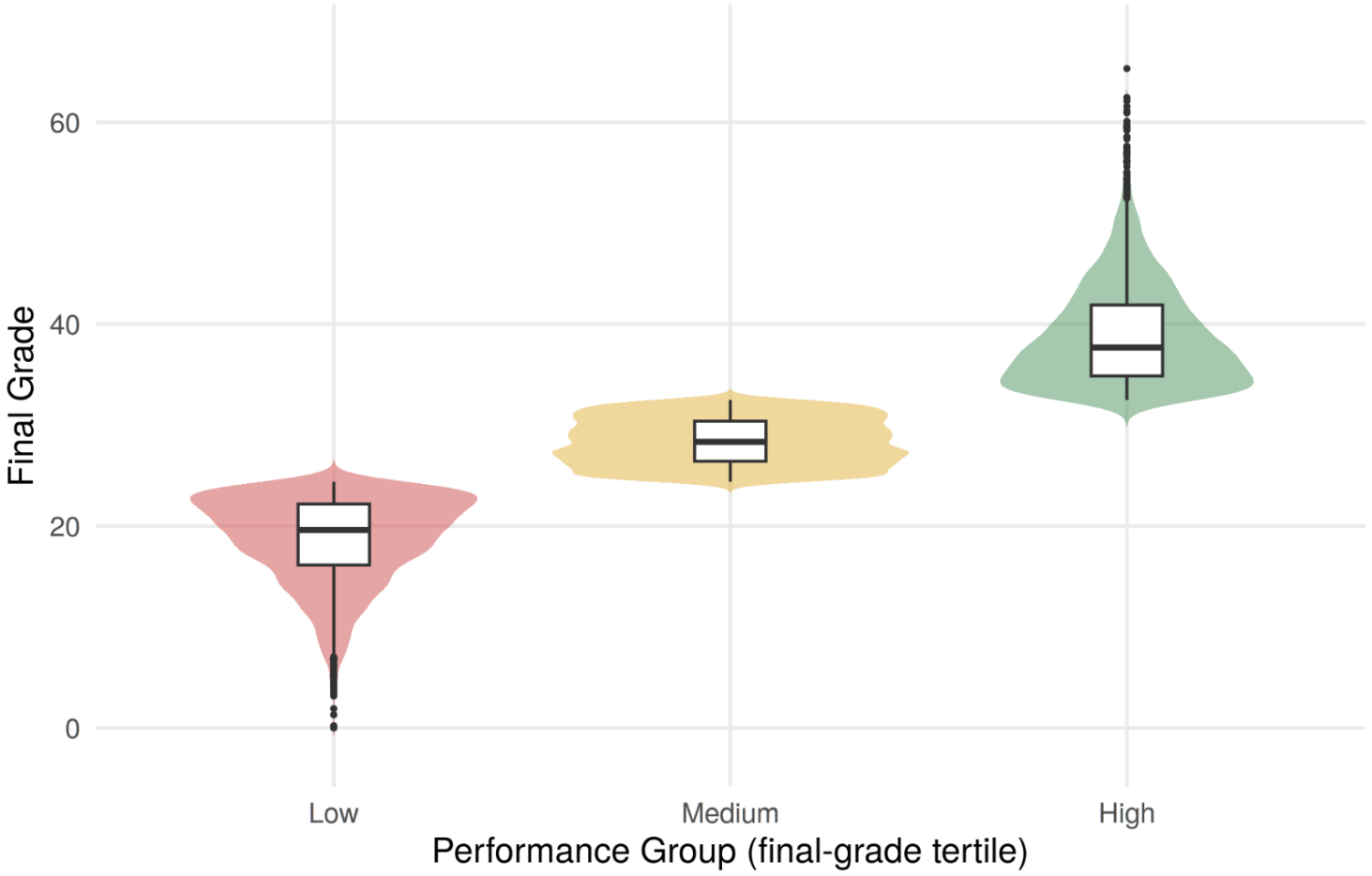


Figure 4

Final grade distributions across derived performance tertiles

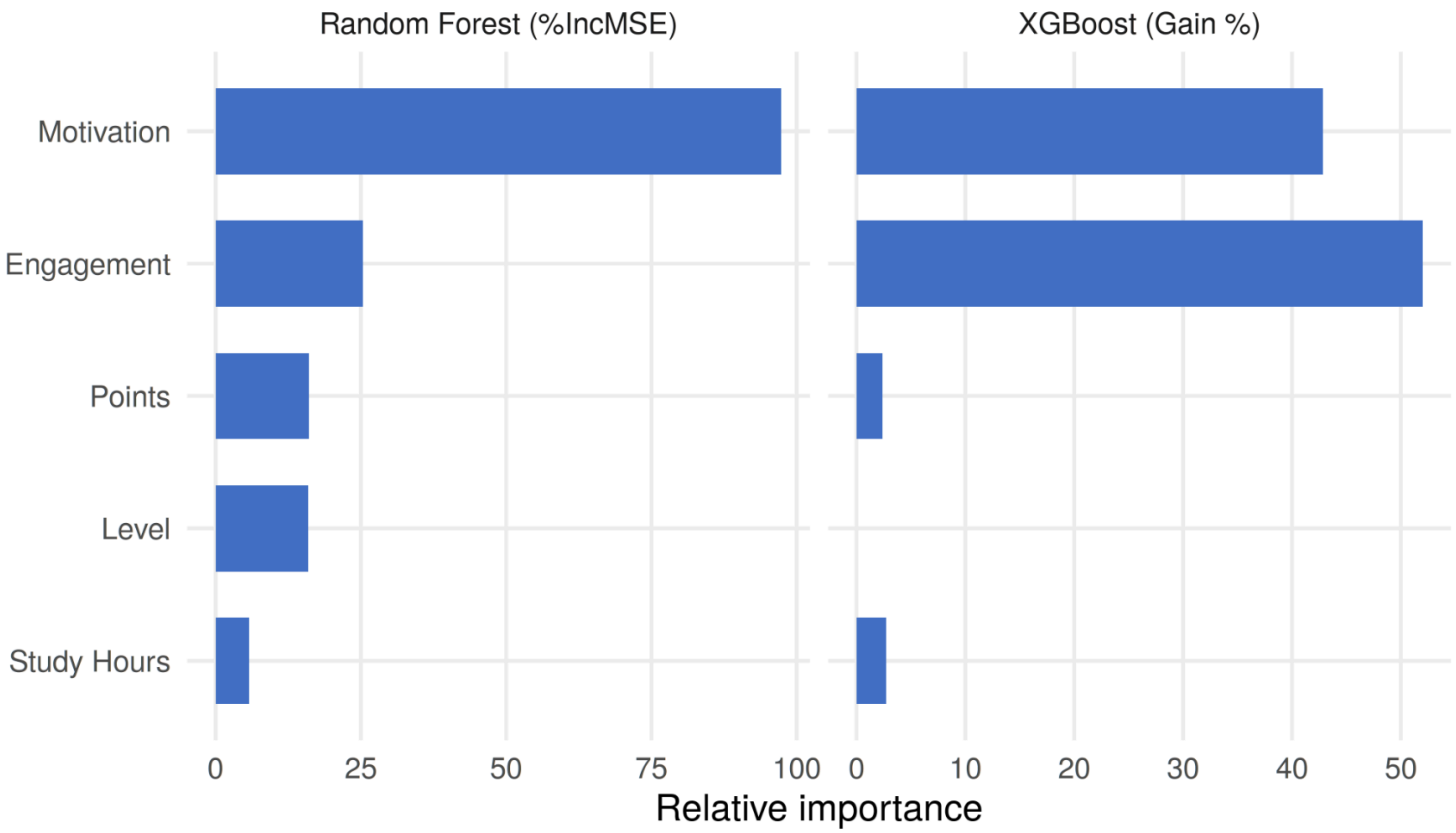


Figure 5

Feature importance for grade prediction: Random Forest permutation importance (%IncMSE) and XGBoost gain

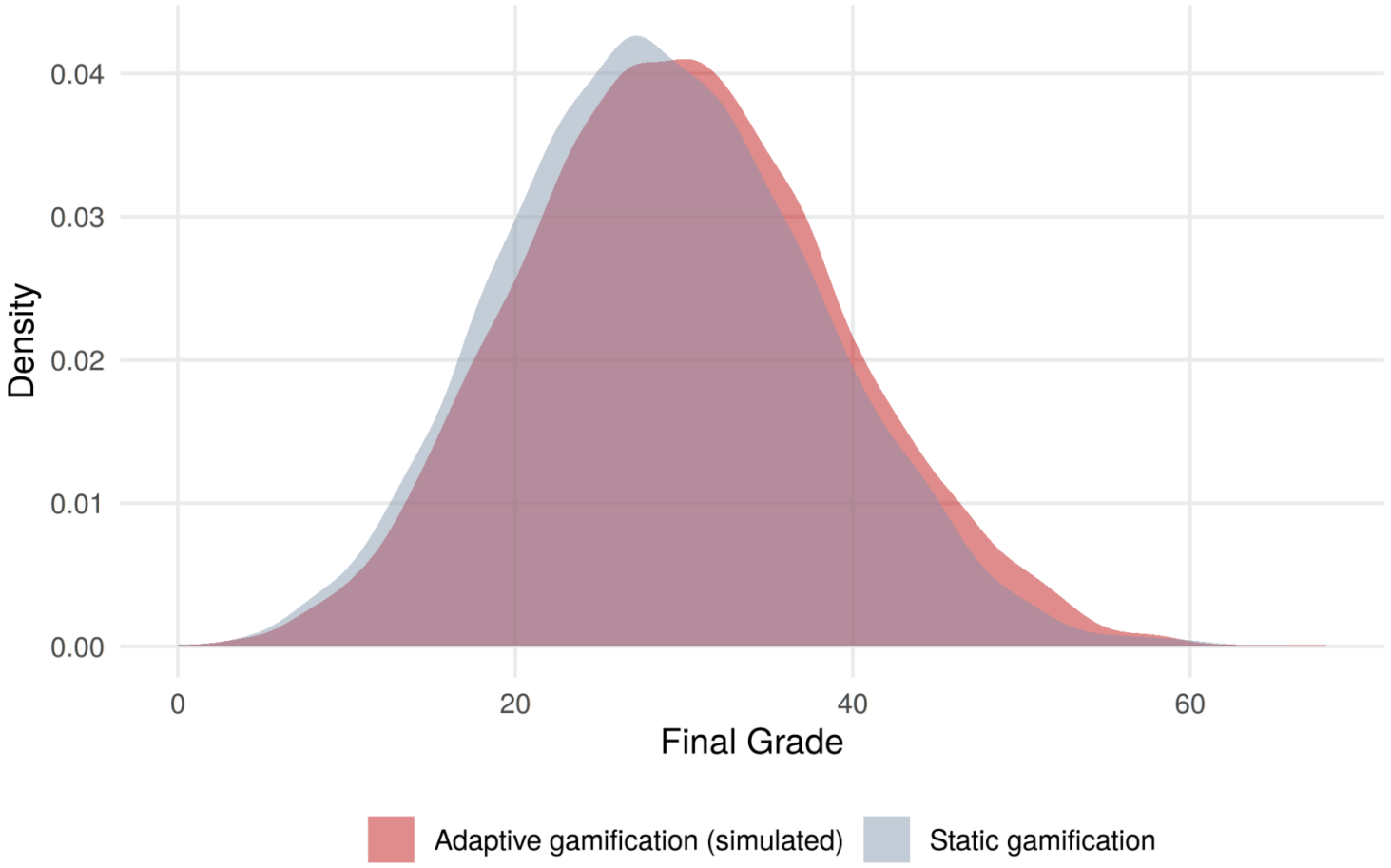


Figure 6

Kernel density of final grades under the observed static condition and the simulated adaptive condition