

Game-Based Learning and Achievement in Undergraduate Statistics and Probability: The Role of Learning Approach

Nurhasniza Idham Abu Hasan

*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,
Perak Branch, Tapah Campus, Malaysia*

ARTICLE INFO

Article history:

Received 07 May 2026
Revised 28 August 2026
Accepted 11 September 2026
Online first
Published 30 September 2026

Keywords:

Game-Based Learning
Learning Approach
Probability Education
Statistics Achievement Undergraduate
Students

DOI:

10.24191/ijelhe.v21n3.21313

ABSTRACT

This study examined undergraduate achievement in statistics and probability under game-based learning (GBL) and traditional instruction, while accounting for baseline achievement and learning approach. A quantitative quasi-experimental pre-test-post-test comparison-group design was used with 305 science-stream undergraduates enrolled in a Malaysian public university. Participants were recruited through convenience sampling and remained in five intact classes, which were randomly allocated at the class level using a random-number generator, with three classes assigned to GBL ($n = 180$) and two classes to traditional instruction ($n = 125$). Achievement was measured using a researcher-developed statistics and probability test, and learning approach was assessed using the R-SPQ-2F. Following a four-week intervention, a 2×2 factorial ANCOVA tested post-test achievement by pedagogical approach and learning approach, with pre-test achievement as the covariate. Pre-test achievement was significantly associated with post-test performance, $F(1, 300) = 117.961$, $p < 0.001$, $\eta_p^2 = 0.282$. After adjustment, pedagogical approach, $F(1, 300) = 59.679$, $p < 0.001$, $\eta_p^2 = 0.166$, and learning approach, $F(1, 300) = 27.200$, $p < 0.001$, $\eta_p^2 = 0.083$, were significant. The Pedagogical Approach $\times$ Learning Approach interaction was also significant, $F(1, 300) = 17.210$, $p < 0.001$, $\eta_p^2 = 0.054$. The GBL-traditional adjusted difference was larger in the Surface Approach group (16.50 points) than in the Deep Approach group (9.20 points). GBL was therefore associated with higher baseline-adjusted achievement. However, the small number of randomly allocated intact classes constrains causal interpretation, and the single-institution design limits generalisability.

*Corresponding author's e-mail address: nurhasniza@uitm.edu.my

1. INTRODUCTION

Statistics and probability are important components of undergraduate education because graduates are increasingly expected to interpret quantitative information, evaluate uncertainty, and make evidence-informed decisions. Competence in these areas requires more than performing calculations. Students must also understand statistical and probabilistic concepts, recognise relationships among them, and apply these concepts appropriately in unfamiliar situations. Accordingly, statistical literacy, reasoning, and thinking remain central goals of statistics education (Batanero et al., 2011; Ben-Zvi & Garfield, 2004).

Despite their importance, statistics and probability remain conceptually challenging for many students. Persistent misconceptions in statistical inference and probability include difficulties with reasoning under uncertainty and interpreting conditional relationships (Castro Sotos et al., 2007; Kaplar et al., 2021; Khazanov & Prado, 2010). Probability learning is particularly demanding because students must interpret uncertain outcomes and relationships rather than reproduce formulas or computational procedures. Recent research similarly highlights the challenge of developing meaningful probabilistic reasoning (Batanero & Álvarez-Arroyo, 2024; Lakrim et al., 2025). These findings suggest that predominantly procedural instruction may not provide sufficient opportunities for students to develop deeper conceptual understanding.

Against this background, game-based learning (GBL) offers an alternative pedagogical approach in which games form an integral part of the learning activity. In GBL, disciplinary content can be embedded within rules, decisions, feedback, and problem-solving tasks. GBL differs from gamification, which involves adding game-design elements to activities that are not themselves games (Deterding et al., 2011; Plass et al., 2015). This distinction is important because GBL engages students directly with disciplinary concepts through the game environment. In probability education, structured games may be particularly relevant because they can involve prediction, repeated trials, outcome comparison, decision-making, and reflection. Such activities may help students connect observable events with formal probability principles. Research in mathematics education has also reported favourable cognitive outcomes associated with GBL, including achievement and mathematical skills (Hui & Mahmud, 2023). However, evidence specific to undergraduate statistics and probability remains limited. Reviews of gamified learning in higher education likewise emphasise variation in how game elements are designed and implemented (Subhash & Cudney, 2018).

A further consideration is students' approach to learning. The Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) distinguishes between the Deep Approach and Surface Approach (Biggs et al., 2001). The Deep Approach reflects a stronger orientation towards meaning and understanding, whereas the Surface Approach is associated with greater reliance on reproductive and task-completion practices. These approaches represent ways of engaging with learning rather than fixed learner types. They may be particularly relevant to GBL because game-based activities often require students to interpret outcomes, compare alternatives, justify decisions, and connect experiences with formal concepts. Students with different learning approaches may therefore respond differently to the same instructional environment, making learning approach a relevant potential moderator of the relationship between pedagogical approach and achievement.

In Malaysian tertiary education, students' understanding of basic statistical concepts has also been examined in relation to teaching and learner-related factors (Ismail et al., 2022). Taken together, the literature identifies a gap at the intersection of pedagogical approach, prior achievement, and learning approach. Although GBL has shown favourable outcomes in broader mathematics education, less is known about its association with achievement in undergraduate statistics and probability after accounting for baseline achievement. It is also unclear whether this association varies according to students' learning approaches. Accordingly, the present study examined post-test achievement under GBL and traditional instruction after controlling for pre-test achievement. It investigated whether learning approach moderated the relationship between pedagogical approach and achievement in a Malaysian public university context.

1.1 Conceptual Framework for GBL, Learning Approach, and Statistics Achievement

The conceptual framework integrates the Foundations of Game-Based Learning with Biggs' cognitive-systems perspective on student learning. Plass et al. (2015) conceptualise educational games as learning environments that can support cognitive, behavioural, affective, and sociocultural engagement, whereas Biggs' Presage-Process-Product model positions learning outcomes as products of interactions between student characteristics, the teaching context, and learning processes (Biggs, 1993). Applied to the present study, pedagogical approach represents the teaching context, initial learning approach represents a measured student characteristic, pre-test achievement represents prior achievement, and post-test achievement in statistics and probability represents the learning product.

Within this framework, GBL is expected to influence achievement by changing how students engage with statistical and probabilistic content. The intervention required active decision-making, repeated encounters with uncertain outcomes, strategy comparison, peer discussion, feedback, and lecturer-guided reflection. These features are theoretically relevant to probability learning because successful performance requires students to interpret uncertainty, connect representations with formal rules, and justify procedures rather than rely only on memorised calculations (Batanero & Álvarez-Arroyo, 2024; Kaplar et al., 2021). Evidence from mathematics education also indicates generally favourable cognitive outcomes associated with well-designed GBL (Hui & Mahmud, 2023). However, this study did not measure decision-making, engagement, feedback uptake, or conceptual processing independently. They are therefore treated as theoretically proposed mechanisms rather than tested mediators.

Learning approach occupies a different role in the framework, as shown in Figure 1. The R-SPQ-2F distinguishes a Deep Approach, characterised by a stronger orientation towards meaning and understanding, from a Surface Approach, characterised by greater reliance on reproductive and task-completion practices (Biggs et al., 2001). Because the R-SPQ-2F was administered before the intervention, learning approach is treated here as an initial moderator rather than an outcome or mediator of GBL. Students entering the same instructional environment may therefore respond differently to its demands. Students classified in the Deep Approach group may already be oriented towards conceptual processing. In contrast, structured GBL may provide additional external prompts for students classified in the Surface Approach group to predict, compare, justify, discuss, and reflect. This creates a theoretically plausible basis for a Pedagogical Approach $\times$ Learning Approach interaction, while leaving the direction and magnitude of that interaction to empirical testing.

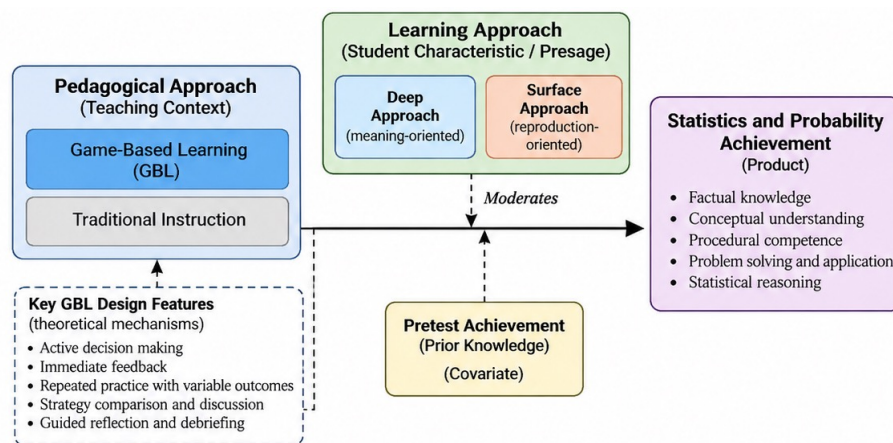


Figure 1 Conceptual framework for GBL, initial learning approach, and statistics and probability achievement
Source: Author (2026)

Note. Solid paths represent relationships examined in the factorial ANCOVA. Initial learning approach is positioned as a moderator of the pedagogical-approach association with post-test achievement, and pre-test achievement is included as a covariate. The GBL design features shown in the figure are theoretically proposed learning mechanisms and were not tested as mediators.

Guided by this framework, the study examined undergraduate achievement in statistics and probability under GBL and traditional instruction while controlling for pre-test achievement. It addressed three questions: (1) Does adjusted post-test achievement differ between GBL and traditional instruction? (2) Does adjusted post-test achievement differ between students initially classified in the Deep Approach and Surface Approach groups? (3) Does initial learning approach moderate the relationship between pedagogical approach and post-test achievement after controlling for pre-test achievement?

2. METHODS AND MATERIALS

2.1 Research Design

This study used a quantitative quasi-experimental pre-test-post-test comparison-group design within an existing undergraduate statistics and probability course. Post-test achievement was the primary outcome, pre-test achievement was the covariate, and pedagogical approach and learning approach were the two between-participant factors. Researchers measured learning approach rather than experimentally manipulating it. The primary model was a 2×2 factorial analysis of covariance (ANCOVA), with GBL versus traditional instruction and Deep Approach versus Surface Approach as the factors.

The study was conducted at a Malaysian public university during the March-to-October semester of the 2025 academic session. The final sample comprised 305 science-stream undergraduates: 180 in the GBL condition and 125 in the traditional-instruction condition. Participants were recruited from among students enrolled in the course. Existing classes remained intact instructional groups, and students were not individually randomised. The five intact classes were randomly allocated to instructional conditions using a random-number generator, with three classes assigned to the GBL condition and two classes assigned to traditional instruction. Students therefore received the instructional condition assigned to their existing class. Both conditions followed the same course syllabus, were taught by the same lecturer, and used comparable instructional and assessment schedules. Random allocation at the class level reduced systematic assignment bias between instructional conditions, although the small number of classes meant that chance class-level imbalance could not be ruled out.

Eligibility for the analytical sample required completion of the pre-test achievement measure, the post-test achievement measure, and the R-SPQ-2F. Of 312 students initially eligible for participation, seven were excluded for not completing all required measures, resulting in a final analytical sample of 305 students. Participants were informed about the study and provided consent before their data were analysed.

2.2 Procedure and Instructional Intervention

Before the intervention, students completed the achievement pre-test and the R-SPQ-2F. The R-SPQ-2F was administered electronically using Google Forms. The intervention then ran for four consecutive weeks. Both conditions covered set theory, counting rules, permutations and combinations, basic probability, addition and multiplication rules, conditional probability, and Bayes' theorem. The learning outcomes required students to solve probability problems, construct representations such as tree diagrams, Venn diagrams, contingency tables, and probability tables, and apply appropriate probability rules.

Traditional instruction relied primarily on lecturer explanation, worked examples, question-and-answer activities, and written exercises followed by class review. In the GBL condition, game activities were integral to instruction rather than conventional activities being supplemented with gamification elements such as points, badges, or rewards. Each GBL activity followed a common instructional sequence: introducing learning objectives and game rules, small-group participation, discussing outcomes and solution strategies, and lecturer-guided reflection. The activities were aligned with the probability content and learning outcomes addressed during the four-week intervention. Table 1 summarises the instructional content, GBL activities, student learning tasks, and intended cognitive processes.

Table 1 *Structure of the Game-Based Learning Intervention*

Week	Probability content	GBL Activity	Student Learning Task	Intended Cognitive Process
1	Set theory and counting rules	Set and Card Classification Challenge	Classify events into appropriate sets, construct relevant set representations, and determine the number of possible outcomes using appropriate counting rules.	Understand, Apply
2	Permutations and combinations	Arrangement and Selection Challenge	Determine whether permutation or combination principles are appropriate under different constraints, calculate possible arrangements or selections, and justify the selected method.	Apply, Analyse
3	Basic probability, addition and multiplication rules	Dice and Card Probability Game	Predict possible outcomes, generate and record experimental results, calculate relevant probabilities, and compare observed outcomes with theoretical probabilities.	Apply, Analyse
4	Conditional probability and Bayes' theorem	Conditional Probability Decision Game	Calculate conditional probabilities, update probabilities as new information becomes available, and justify probability-based decisions using Bayes' theorem.	Apply, Analyse, Evaluate

For descriptive purposes, the cognitive demands of the GBL activities were mapped onto the revised Bloom's taxonomy (Anderson & Krathwohl, 2001). The activities progressed from understanding and application towards analysis and evaluation as students moved from classifying outcomes to comparing strategies and making probability-based decisions.

2.3 Measures

2.3.1 Statistics and Probability Achievement Test

Achievement was assessed using a researcher-developed statistics and probability test aligned with the course learning outcomes. The test contained three structured-response questions with 10 sub-items and a possible score ranging from 0 to 100. Items addressed set theory and counting rules, permutations and combinations, probability rules, conditional probability, and Bayes' theorem. A predetermined analytical rubric awarded marks for identifying relevant concepts, selecting procedures, completing calculations, constructing appropriate representations where required, and interpreting results. For example, one conditional-probability item allocated two marks for identifying the conditional event, three for selecting and applying the rule, three for calculation, and two for contextual interpretation. Equivalent pre-test and post-test forms were used to assess the same learning outcomes, with variations in numerical values and problem contexts introduced to minimise recall effects and reduce measurement bias in comparisons between assessed the same learning outcomes, with variations in numerical values and problem contexts to minimise recall effects and reduce measurement bias when comparing the two pedagogical conditions.

The achievement test was reviewed by three experts in statistics education, mathematics education, and educational measurement. Three experts in statistics education, mathematics education, and educational measurement reviewed the achievement test. The review focused on content coverage, alignment with the learning outcomes, clarity, difficulty, and suitability for undergraduate students. A pilot study involving 30 students who were of 30 students not included in the main analytical sample examined item clarity, administration procedures, test duration, and the usability of the marking

rubric marking rubric's usability. Internal consistency was evaluated using Cronbach's alpha because the structured-response components used multiple score categories. The pilot administration produced a coefficient of $\alpha = 0.81$, indicating acceptable internal consistency. In the main sample, the corresponding coefficient was $\alpha = 0.84$.

Responses were scored using a detailed predetermined marking rubric. Because all responses were scored by a single trained marker, a single trained marker scored all responses, inter-rater reliability was not estimated. To enhance scoring consistency, the same marking criteria were applied to all scripts, and ambiguous responses were reviewed with reference to improve scoring consistency, the same marking criteria were applied to all scripts, and ambiguous responses were reviewed against the predetermined scoring guide.

2.3.2 Revised Two-Factor Study Process Questionnaire

Learning approach was assessed using the 20-item R-SPQ-2F developed by Biggs et al. (2001). Ten items contribute to the Deep Approach scale and ten to the Surface Approach scale. Items were rated on a five-point response scale from 1 (never or only rarely true of me) to 5 (always or almost always true of me). Item scores were summed separately, producing scale scores from 10 to 50, with higher values indicating stronger endorsement of the corresponding approach. For the factorial ANCOVA, the study created mutually exclusive categories from the relative scale scores. Students with Deep Approach scores greater than Surface Approach scores were classified in the Deep Approach group, and those with Surface Approach scores greater than Deep Approach scores were classified in the Surface Approach group. Cases with equal scores were handled using the higher motive subscale score. The pilot study involving 30 students also examined R-SPQ-2F item comprehension in the study context. Internal consistency was evaluated separately for the two principal R-SPQ-2F dimensions using Cronbach's alpha. In the pilot sample, the Deep Approach scale yielded $\alpha = 0.82$, while the Surface Approach scale yielded $\alpha = 0.79$. In the main sample, the corresponding coefficients were $\alpha = 0.85$ and $\alpha = 0.81$, respectively.

2.4 Data Analysis

Data were analysed using IBM SPSS Statistics version 26.0, with alpha set at .05. Descriptive statistics summarised pre-test and post-test achievement. Data screening included checks for missing data and standardised residuals. The 2×2 factorial ANCOVA model used post-test achievement as the dependent variable, pedagogical approach and learning approach as between-participant factors, and pre-test achievement as the covariate. Homogeneity of variance was evaluated with Levene's test. Homogeneity of regression slopes was assessed through interactions between the pre-test covariate and the grouping factors. Linearity, homoscedasticity, and approximate residual normality were assessed using residual-versus-predicted plots and a normal Q-Q plot. Main effects and interactions were reported with F statistics, degrees of freedom, p values, and partial eta squared (η^2), η_p^2 . A significant Pedagogical Approach $\times$ Learning Approach interaction was followed by estimated marginal means and simple-effects comparisons with mean differences, standard errors, p values, and 95% confidence intervals. Pairwise simple-effects comparisons were conducted separately within each level of the conditioning factor and adjusted using the Bonferroni procedure within each comparison family. Simple-effects tests used the residual error term from the ANCOVA model.

3. RESULTS

3.1 Descriptive Results and Assumption Checks

The final sample included 305 students. Table 2 shows the descriptive achievement scores by pedagogical approach and learning approach. Within each learning-approach group, pre-test means were close across instructional conditions. Post-test means were higher in the GBL condition in both groups, with a larger descriptive separation in the Surface Approach group.

Table 2 Results of Pre-test and Post-test Achievement by Pedagogical Approach and Learning Approach

Learning approach	Condition	n	Pre-test M (SD)	Post-test M (SD)
Deep Approach	GBL	94	68.20 (8.60)	85.10 (8.40)
Deep Approach	Traditional instruction	65	67.80 (8.50)	76.00 (9.20)
Surface Approach				
GBL	86	63.40 (9.00)	78.30 (8.90)	
Surface Approach				
Traditional instruction	60	63.10 (9.10)	62.40 (9.70)	

Note. Total pre-test and post-test scores ranged from 0 to 100.

3.2 Preliminary Analysis and Statistical Assumptions

The assumption checks presented in Table 3 indicated no serious violations of the ANCOVA assumptions. Levene's test was non-significant, $F(3, 301) = 1.12$, $p = 0.341$, indicating that the homogeneity of variance assumption was met.

Table 3 Results of Assumption Tests for the Factorial ANCOVA

Assumption	Test	Statistic	p-value	Decision
Homogeneity of variance	Levene's test	$F(3, 301) = 1.12$	0.341	Met
Homogeneity of regression slopes	Pre-test $\times$ Pedagogy	$F(1, 297) = 1.34$	0.248	Met
Homogeneity of regression slopes	Pre-test $\times$ Learning Approach	$F(1, 297) = 0.76$	0.384	Met
Homogeneity of regression slopes	Pre-test $\times$ Pedagogy $\times$ Learning Approach	$F(1, 297) = 0.41$	0.522	Met

The pre-test $\times$ pedagogical approach interaction, $F(1, 297) = 1.34$, $p = 0.248$, the pre-test $\times$ learning approach interaction, $F(1, 297) = 0.76$, $p = 0.384$, and the pre-test $\times$ pedagogical approach $\times$ learning approach interaction, $F(1, 297) = 0.41$, $p = 0.522$, were also non-significant. These results provided no evidence that the relationship between pre-test and post-test achievement differed across the study conditions, supporting the homogeneity of regression slopes assumption. Visual inspection of the standardised residuals in Figure 2 revealed no systematic pattern across the range of predicted values, indicating no substantial departures from linearity or homoscedasticity. The normal Q-Q plot indicated approximately normal residuals.

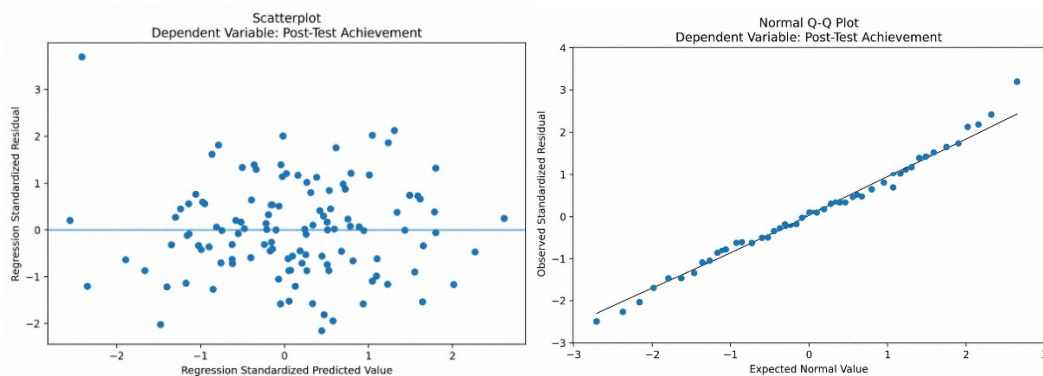


Figure 2 Diagnostic plots for the factorial ANCOVA model
Source: Author (2026)

3.3 Factorial ANCOVA and Interaction

Table 4 presents the factorial ANCOVA. Pre-test achievement was significantly associated with post-test performance, $F(1, 300) = 117.961$, $p < 0.001$, $\eta_p^2 = 0.282$. This was the largest partial eta-squared value in the model, indicating that prior achievement remained an important source of variation in post-test performance and supporting its inclusion as a covariate.

Table 4 Results of Factorial ANCOVA for Post-test Achievement

Source	SS	df	MS	F	p	η_p^2
Pre-test achievement	4250.310	1	4250.310	117.961	< .001	.282
Pedagogical approach	2150.450	1	2150.450	59.679	< .001	.166
Learning approach	980.120	1	980.120	27.200	< .001	.083
Pedagogical approach $\times$ Learning approach	620.150	1	620.150	17.210	< .001	.054
Error	10810.200	300	36.034			
Corrected Total	18811.230	304				

After adjustment for pre-test achievement, pedagogical approach was significant, $F(1, 300) = 59.679$, $p < 0.001$, $\eta_p^2 = 0.166$. Learning approach was also significant, $F(1, 300) = 27.200$, $p < 0.001$, $\eta_p^2 = 0.083$. These main effects, along with the significant Pedagogical Approach $\times$ Learning Approach interaction, $F(1, 300) = 17.210$, $p < 0.001$, $\eta_p^2 = 0.054$, indicate that the η_p^2 difference between GBL and traditional instruction varied by learning approach.

Estimated marginal means clarified this interaction. In the Deep Approach group, adjusted post-test achievement was 85.40 (SE = 1.85) under GBL and 76.20 (SE = 1.88) under traditional instruction. In the Surface Approach group, the corresponding adjusted means were 78.60 (SE = 1.90) and 62.10 (SE = 1.92). The adjusted GBL-traditional difference was therefore 9.20 points in the Deep Approach group and 16.50 points in the Surface Approach group. This pattern indicates that the relative advantage associated with game-based learning was more pronounced among students classified in the Surface Approach group. Figure 3 illustrates the interaction pattern.

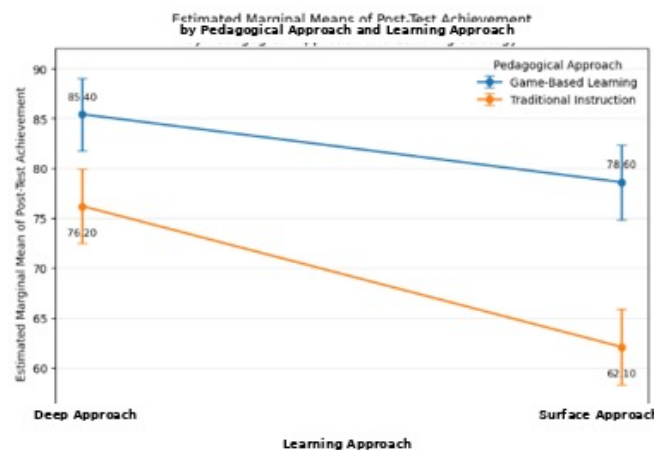


Figure 3 Estimated marginal means of post-test achievement by pedagogical approach and learning approach after adjustment for pre-test achievement. Note. Error bars represent 95% confidence intervals.

Source: Author (2026)

3.4 Simple-Effects Comparisons

Table 5 shows simple-effects comparisons. GBL was associated with higher adjusted achievement within both learning-approach groups. However, the adjusted GBL-traditional difference was larger in the Surface Approach group (16.50 points) than in the Deep Approach group (9.20 points). All simple-effects tests used 300 residual degrees of freedom from the ANCOVA model. Bonferroni adjustment was specified separately within each simple-effects family.

Table 3 Results of Assumption Tests for the Factorial ANCOVA

Comparison	Mean difference	SE	t(300)	p	95% CI
GBL vs traditional instruction, Deep Approach	9.20	2.64	3.485	< 0.001	[3.97, 14.43]
GBL vs traditional instruction, Surface Approach	16.50	2.67	6.180	< 0.001	[11.21, 21.79]
Deep vs Surface Approach, GBL	6.80	2.65	2.566	0.011	[1.55, 12.05]
Deep vs Surface Approach, traditional instruction	14.10	2.68	5.261	< 0.001	[8.79, 19.41]

The Deep-Surface difference was also smaller under GBL (6.80 points) than under traditional instruction (14.10 points). Accordingly, the achievement gap associated with learning approach was 7.30 points smaller in the GBL condition. This does not mean that GBL eliminated learning-approach differences because the Deep Approach group still had higher adjusted achievement within the GBL condition.

4. DISCUSSION

This study examined undergraduate achievement in statistics and probability under game-based learning (GBL) and traditional instruction after controlling for pre-test achievement and considering initial learning approach. Three findings are central to the discussion. GBL was associated with higher adjusted post-test achievement than traditional instruction; students in the Deep Approach group achieved higher adjusted scores than those in the Surface Approach group; and the association between pedagogical approach and achievement varied by learning approach.

Pre-test achievement remained an important source of variation in post-test performance. Its strong association with post-test achievement highlights the importance of accounting for prior performance when comparing instructional conditions in a quasi-experimental study. Including pre-test achievement as a covariate strengthened the comparison by adjusting for observed baseline differences. However, ANCOVA does not establish baseline equivalence or remove unmeasured differences between intact classes. The findings should therefore be interpreted as baseline-adjusted associations rather than definitive causal effects of instructional condition.

4.1 Pedagogical approach and adjusted achievement

After adjustment for pre-test achievement, GBL was associated with higher post-test achievement than traditional instruction, indicating that measured baseline achievement alone did not account for the observed difference between the instructional conditions. This pattern is broadly consistent with previous research. Tokac et al. (2019) reported positive mathematics-achievement outcomes associated with game-based learning, while Hui and Mahmud (2023) found generally favourable cognitive outcomes in mathematics education. Reviews from other higher-education contexts also report learning benefits from structured game-based or serious-game interventions (Kuruca Ozdemir & Dinc, 2022; Platz, 2022; Sierra-Daza et al., 2024). The present study extends this evidence to undergraduate statistics and probability while explicitly adjusting for baseline achievement.

One plausible explanation lies in the structure of the GBL activities. Students classified events, selected probability principles, predicted and compared outcomes, justified decisions, discussed strategies, and participated in lecturer-guided reflection. These activities may have provided repeated opportunities to connect game experiences with formal probability concepts rather than rely mainly on procedural reproduction. This interpretation is consistent with Plass et al. (2015), who emphasised that the educational value of games depends on the learning processes embedded within the game environment. Because these processes were not measured directly, they should be regarded as plausible explanations rather than demonstrated mediators.

4.2 Learning Approach and Adjusted Achievement

Learning approach was also associated with adjusted post-test achievement. Students classified in the Deep Approach group achieved higher adjusted scores than those classified in the Surface Approach group. This finding is consistent with the R-SPQ-2F framework, which distinguishes meaning-oriented engagement from more reproductive and task-completion practices (Biggs et al., 2001). Recent systematic evidence has likewise linked learning approaches with academic performance across higher-education contexts (Roehe et al., 2024). In statistics and probability, a stronger orientation towards understanding relationships may be advantageous because successful performance requires students to identify relevant concepts, select procedures, interpret representations, and explain results. Nevertheless, Deep and Surface Approaches represent measured approaches to learning rather than fixed learner identities.

4.3 Initial Learning Approach as a Moderator

The Pedagogical Approach $\times$ Learning Approach interaction was statistically significant, showing that the association between pedagogical approach and achievement was not uniform across learning-approach groups. The adjusted GBL-traditional difference was 9.20 points among students classified in the Deep Approach group and 16.50 points among those in the Surface Approach group. The Deep-Surface difference was also smaller under GBL (6.80 points) than under traditional instruction (14.10 points). Thus, although the Deep Approach group achieved higher scores overall, the GBL difference was larger in the Surface Approach group.

A plausible interpretation is that students classified in the Deep Approach group may already engage more readily in meaning-oriented processing across instructional settings. By contrast, structured GBL may provide additional external prompts for students classified in the Surface Approach group to predict, compare, justify, discuss, and reconsider their reasoning. Such demands could help explain the larger adjusted GBL-traditional difference in this group. However, the study did not directly measure these learning processes. The findings also do not demonstrate that GBL changed students' underlying learning approaches. The appropriate interpretation is that initial learning approach moderated the association between pedagogical approach and adjusted achievement.

4.4 Instructional Implications

From an instructional perspective, the findings suggest that the value of GBL depends on how game activities align with learning outcomes and integrate with formal teaching. Structured games may be most useful when they require students to predict outcomes, apply probability rules, compare solutions, justify decisions, and connect game experiences with formal concepts through lecturer-guided reflection. GBL should therefore complement, rather than replace, explanation, worked examples, and written practice. The interaction also suggests that the same instructional environment may not produce the same achievement differences across students with different learning approaches.

4.5 Limitations and Future Research

Several limitations should be considered. Although the five intact classes were randomly allocated to instructional conditions, the small number of classes limits how well randomisation can balance class-level characteristics. The reported ANCOVA was conducted at the individual-student level and did not explicitly model classroom clustering, so residual class-level differences may have contributed to the observed findings. The four-week intervention, the science-stream sample, and the single Malaysian public university context further limit generalisability. In addition, the achievement assessment was researcher-developed and scored by a single trained marker. Categorising the continuous R-SPQ-2F dimensions into mutually exclusive groups also reduced the information retained from the original scale scores.

Future research should replicate the study across multiple institutions and larger numbers of intact classes, using larger cluster-randomised designs and, where feasible, cluster-aware statistical models. Longer interventions would help determine whether the observed differences persist over time. Future studies should also consider modelling Deep Approach and Surface Approach scores continuously and administering the R-SPQ-2F longitudinally to distinguish moderation by initial learning approach from changes in learning approach. Direct measures of conceptual reasoning, feedback use, engagement, strategy revision, and peer interaction would further clarify the processes that may underlie the larger adjusted GBL-traditional difference observed among students classified in the Surface Approach group.

5. CONCLUSION

This study found that GBL was associated with higher post-test achievement in undergraduate statistics and probability after adjustment for pre-test achievement. Learning approach was also associated with achievement, and the significant Pedagogical Approach $\times$ Learning Approach interaction showed that the adjusted GBL-traditional difference was larger in the Surface Approach group than in the Deep Approach group. The findings suggest that structured games aligned with probability learning outcomes, followed by discussion and guided reflection, may be useful in undergraduate probability instruction. For instructional practice, these findings favour deliberately integrating game tasks with explicit probabilistic reasoning rather than treating games as stand-alone enrichment activities. Such alignment is important if game participation is to remain connected to assessable course learning outcomes. These findings do not show that GBL changed students' underlying learning approaches.

6. ACKNOWLEDGEMENTS

The author expresses sincere gratitude to Universiti Teknologi MARA for the institutional support and facilities provided throughout the study. The author also thanks all participants who generously contributed their time during data collection.

7. FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

8. AUTHORS' CONTRIBUTIONS

Nurhasniza Idham Abu Hasan: Conceptualisation, methodology, formal analysis, investigation, validation, writing the original draft, editing, and final approval of the manuscript.

9. CONFLICT OF INTEREST DECLARATION

I certify that this article is my original work. It has not been published previously, submitted for publication elsewhere, or placed under consideration by another journal. I attest that I have contributed significantly to the work and take responsibility for the validity and legitimacy of the data and their interpretation in this submission to IJELHE.

10. DECLARATION OF GENERATIVE AI IN THE WRITING PROCESS

During the preparation of this work, the author used ChatGPT (OpenAI) to enhance the clarity of the writing. The author subsequently reviewed and edited the content as needed and takes full responsibility for the published content.

11. DATA AVAILABILITY/SUPPLEMENTARY MATERIALS

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

12. REFERENCES

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Batanero, C., & Álvarez-Arroyo, R. (2024). Teaching and learning of probability. *ZDM - Mathematics Education*, 56, 5–17. <https://doi.org/10.1007/s11858-023-01511-5>
- Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics: Challenges for teaching and teacher education: A joint ICMI/IASE study*. Springer. <https://doi.org/10.1007/978-94-007-1131-0>
- Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–15). Springer. https://doi.org/10.1007/1-4020-2278-6_1
- Biggs, J. B. (1993). From theory to practice: A cognitive systems approach. *Higher Education Research & Development*, 12(1), 73–85. <https://doi.org/10.1080/0729436930120107>
- Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71(1), 133–149. <https://doi.org/10.1348/000709901158433>
- Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98–113. <https://doi.org/10.1016/j.edurev.2007.04.001>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). Association for Computing Machinery. <https://doi.org/10.1145/2181037.2181040>
- Hui, H. B., & Mahmud, M. S. (2023). Influence of game-based learning in mathematics education on students' cognitive and affective domains: A systematic review. *Frontiers in Psychology*, 14, Article 1105806. <https://doi.org/10.3389/fpsyg.2023.1105806>
- Ismail, N. Z.-I., Abu Kassim, N. L., & Mahmud, Z. (2022). Factors influencing students' understanding of basic statistical concepts. *IJUM Journal of Educational Studies*, 10(2), 174–204. <https://doi.org/10.31436/ijes.v10i2.459>
- Kaplar, M., Lužanin, Z., & Verbić, S. (2021). Evidence of probability misconception in engineering students: Why even an inaccurate explanation is better than no explanation. *International Journal of STEM Education*, 8, Article 18. <https://doi.org/10.1186/s40594-021-00279-y>
- Khazanov, L., & Prado, L. (2010). Correcting students' misconceptions about probability in an introductory college statistics course. *Adults Learning Mathematics*, 5(1), 23–35. <https://files.eric.ed.gov/fulltext/EJ1068215.pdf>
- Kuruca Ozdemir, E., & Dinc, L. (2022). Game-based learning in undergraduate nursing education: A systematic review of mixed-method studies. *Nurse Education in Practice*, 62, Article 103375. <https://doi.org/10.1016/j.nepr.2022.103375>
- Lakrim, A., Chergui, M., & El Wahbi, B. (2025). An exploratory study of the difficulties related to the probabilistic modelling process. *Journal of Education and Learning (EduLearn)*, 19(3), 1590–1599. <https://doi.org/10.11591/edulearn.v19i3.22046>
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. <https://doi.org/10.1080/00461520.2015.1122533>

Platz, L. (2022). Learning with serious games in economics education: A systematic review of the effectiveness of game-based learning in upper secondary and higher education. *International Journal of Educational Research*, 115, Article 102031. <https://doi.org/10.1016/j.ijer.2022.102031>

Roehe, M. A., Trost, C., Grundnig, J. S., Anvari-Pirsch, A., & Holzinger, A. (2024). Evaluating the dynamics of learning approaches: A systematic review investigating the nexus between teaching methods and academic performance in medical and dental education. *Educational Psychology Review*, 36, Article 118. <https://doi.org/10.1007/s10648-024-09952-4>

Sierra-Daza, M. C., Martín-del-Pozo, M., & Fernández-Sánchez, M. R. (2024). Videojuegos y aprendizaje en la etapa universitaria: Una revisión sistemática. *Revista de Educación*, 405, 103–132. <https://doi.org/10.4438/1988-592X-RE-2024-405-629>

Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behaviour*, 87, 192–206. <https://doi.org/10.1016/j.chb.2018.05.028>

Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).