Journal of Learning Theory and Methodology
Журнал теорії та методології навчання

ISSN 2708-7581 (Online)
ISSN-L 2708-7573

Original Scientific Article

Evaluating Teaching Effectiveness and Student Performance Across Diverse Courses: An Analysis of Final Exam Scores and Teaching Techniques

Vicente E. Montano[1ABCD] and Glenndon C. Sobrejuanite[1ABCD]

[1]University of Mindanao

Authors' Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

DOI: 10.17309/jltm.2025.6.1.03

Abstract

Objectives. This study examines teaching effectiveness and student performance for four business-related courses: Monetary Policy and Central Banking (FM 221), Good Governance and Social Responsibility (BAHR 213), Consumer Behavior (MM 212), and Introduction to Microeconomics (BE 121).

Materials and methods. 147 students participated in the study, and Bayesian pairwise comparison, descriptive statistics, and effect size analysis were used to determine which courses had significantly different performance scores.

Results. The results show that BAHR 213 and MM 212 students performed better than FM 221 and BE 121 students, indicating the role of active learning strategies, structured content delivery, and student engagement techniques for better learning. Lower performance with higher variability of scores in FM 221 and BE 121 indicates the requirement for pedagogical improvements, additional support for students, and curriculum modifications.

Conclusions. Bayesian analysis confirmed that the differences are statistically significant: the first discriminant function explains 86.8% of the variation, $p < 0.001$; the second function explains 13.2%, $p < 0.001$. The paper concludes with a discussion on the need to refine instructional methods and targeted interventions to improve student outcomes. Future studies need to look into longitudinal studies and controlled pedagogical experiments that can further validate these findings and enhance best practices for effective teaching strategies.

Keywords: teaching effectiveness, student performance, Bayesian analysis, pedagogical strategies, business education, learning outcomes.

Introduction

There are various challenges in appraising teaching effectiveness and students' performance across different courses, primarily when evaluated based on scores from final exams and teaching techniques (Akiri, 2013). This happens because course content, demographic characteristics of students, and methodologies of instruction vary, complicating the assessment process (Phye, 1984).

The biggest challenge is variability in the content and structure of courses. Different courses have varying learning objectives, complexity levels, and assessment methods, and using a standardized evaluation for each course is rather complicated (Emery et al., 2003). For instance, a Monetary Policy and Central Banking course focuses intensely on intricate economic theory, whereas Consumer Behavior is taught by application and case studies. This diversity requires specific assessment criteria that can better capture the individual demands of each course (Uttl et al., 2017).

A key issue is that student demographics are diverse. These students differ in prior knowledge, learning style, and circumstance, which may affect class performance (Krätzig & Arbuthnott, 2006). For instance, students with rich exposure to economics are more likely to find courses like Introduction to Microeconomics easy than those with minimal exposure to this subject. An assessment tool for teaching efficacy has to accommodate these differences, making assessment fair and effective (Kleinsasser, 2014).

The effectiveness of teaching techniques varies widely among instructors, even within the same course. Different instructors employ different pedagogical approaches, such as lecture-based, active, or blended learning (Delialioğlu, 2012). These variations impact student engagement and learning outcomes. Additionally, instructors' experience, expertise, and teaching style influence their effectiveness. Evaluating teaching techniques requires a nuanced approach, considering these individual differences and their impact on student performance (Koopman & Beijaard, 2024).

Dependence on final exam grades as the significant tool for assessing teaching effectiveness and student performance

Corresponding Author: Vicente Montano,
e-mail: vicente_montano@umindanao.edu.ph

has weaknesses. Final exams usually measure a narrow scope of skills and knowledge, typically involving memorization and problem-solving within a time frame (Darling-Hammond et al., 2010). This approach may only reflect part of the scope of student learning and development, such as critical thinking, creativity, and practical application of knowledge (Birenbaum & Dochy (Eds.), 2012). In addition, performance on the exam can be influenced by test anxiety and other extraneous pressures that do not necessarily reflect the student's proper understanding or the instructor's effectiveness (Brookhart, 2011).

Another challenge is the validity and reliability of methods of evaluation. Validity is the degree to which an assessment measures what it is supposed to measure. Reliability refers to the consistency of assessment results (Ratanawongsa et al., 2008). Designing evaluation tools that are valid and reliable in diverse courses is, therefore, demanding and requires careful design with constant refinement. This is in terms of the alignment of assessments with learning objectives, the use of multiple measures of performance, and the periodic review and updating of criteria for evaluation (Price, 2015).

Bias and subjectivity are challenges in teaching effectiveness and student performance assessment. For instance, the biases of the instructors, the perceptions of students, and priorities at an institutional level affect evaluation (Tripon, 2019). For example, an instructor's likability and difficulty in a course may impact teaching performance while being evaluated by the students. Quantitative and qualitative data may address these biases, such as peer review, self-assessment, and objective performance measures (Whitesides & Beck, 2020).

Evaluating teaching effectiveness and student performance in various courses requires a broad, multi-faceted approach. It involves course content variability, student demographics, instructional methods, validity, reliability, and fairness of the assessment tools (Carpenter et al., 2020). By considering such challenges and adopting best practices in educational assessment, institutions will better understand and enhance the quality of teaching and learning (Chetty et al., 2014).

Theoretical Framework

The theoretical framework of this study is based on educational theories that emphasize the relationship between teaching techniques, course characteristics, and student performance. This framework gives a structured approach to understanding how different variables interact to influence learning outcomes.

The study draws from constructivist theories of learning, which advocate for students' active engagement and participation in the learning process (Hein, 1991). Techniques like active learning, formative assessments, and support resources are key to this approach. Active learning is a strategy that engages students in processing and applying information actively, such as group discussions, problem-solving activities, and case studies (Kumar, 2013). Formative assessments, which include quizzes and assignments, are continuous feedback that helps students know their strengths and weaknesses (Clark, 2012). Support resources like tutoring and study groups also assist students who require such resources to ensure that every student is given a fair chance at success (Hagstrom, 2006).

It considers the courses' characteristics, including the content's complexity, difficulty level, and methods to engage the students. These aspects help determine how well the students understand and retain what has been taught (Duane & Satre, 2014). Courses that are well-structured and aligned with students' existing knowledge and learning ability will likely be high performers. The study analyses these characteristics across the four courses and discusses how they impact student performance (Collins, 2008).

Performance is gauged based on final examination scores and the extent to which students comprehend and master course material. Such variability and consistency in scores contribute further to the knowledge of effectiveness, the teaching methods applied, and the characteristics of the course (Dwiyanti, 2024, June). High variability in scores reflects that some students are having issues with others, meaning students need more focused support and perhaps changes in teaching strategy (Wu et al., 2004, January).

The result of the theoretical framework is excellence in teaching and learning. The study aims to provide actionable recommendations for improving teaching practices by identifying factors that explain differences in student performance. This involves adopting more effective teaching techniques, revising course content, and providing additional student support. The framework emphasizes the importance of continuous professional development and improvement for faculty, so all students can fully develop their abilities.

The study's theoretical framework integrates education theories into practical teaching and learning considerations. The framework enables a comprehensive understanding of how interacting variables influence students' performances and acts as a basis to formulate well-informed recommendations for improvements in teaching effectiveness and learning results.

Both teaching techniques and course characteristics determine student performance. Student performance determines the study's outcome, which encompasses excellence in teaching and learning and recommendations for improvement (Aydogdu & Ay, 2016). This framework helps understand how various teaching methods and course characteristics determine student performance and how it contributes to the quality of teaching and learning (Nurhuda et al., 2023).

By identifying the factors contributing to differences in student performance, the study aims to give actionable recommendations for teaching improvement. This includes using more effective teaching techniques (Fernando & Marikar, 2017), revising course content (Moallem, 2001), and providing additional student support (Ng'ambi & Johnston, 2006). The framework emphasizes the importance of continuous improvement and professional development for faculty to ensure that all students have the opportunity to reach their full potential.

Therefore, this theoretical framework of the research study integrates theoretical aspects of education with practical teaching and learning considerations. Thus, this study provides a more holistic approach toward understanding how different variables may interact to influence student performance outcomes and allows a basis for making recommendations informed by the findings on improving teaching effectiveness and learning.

This research is aligned with the United Nations Sustainable Development Goal 4: Quality Education. By comparing the student's performance in different courses and identifying areas of disparity, this research study focuses on educational interventions to ensure equitable learning opportunities. This will result in better teaching methods, improved student support, and more effective curricula for enhancing the quality of education. This aligns with the principle of ensuring inclusive and equitable quality education by promoting lifelong learning opportunities for everyone.

The objective of this study is to assess the effectiveness of teaching methods and student performance in four different courses: Monetary Policy and Central Banking (FM 221), Good Governance and Social Responsibility (BAHR 213), Consumer Behavior (MM 212), and Introduction to Microeconomics (BE 121) through final exam scores. The study intends to identify significant differences in student performance, understand the effects of various teaching techniques, and provide actionable recommendations for improving teaching practices. This study will examine the variability and consistency of exam scores and conduct statistical analyses to shed light on how different approaches to instruction affect learning and promote excellence in teaching and learning.

Materials and Methods

Participants

This research was conducted with the involvement of 147 students enrolled in four different courses offered in a higher education institution, namely, Monetary Policy and Central Banking (FM 221), Good Governance and Social Responsibility (BAHR 213), Consumer Behavior (MM 212), and Introduction to Microeconomics (BE 121), all selected based on their registration for these classes during the First Semester of the SY 2024-2025. The study focused on the student's final exam scores to evaluate the effectiveness of different teaching methods. The students were from diverse academic backgrounds in business, with varying levels of prior knowledge and engagement in economics, governance, and business studies. The sample was designed to capture a range of student performance levels, allowing for a comparative analysis of teaching strategies and their impact on learning outcomes. All student data was kept confidential and used only for research purposes, thus upholding ethical considerations.

Methods of Research

The research followed a systematic approach, using statistical techniques, data collection procedures, and pedagogical testing to ensure that teaching effectiveness was scientifically evaluated. The primary focus was analyzing students' final exam scores to identify performance variations across courses and infer the influence of different instructional methods. The study integrated quantitative and inferential statistical techniques to provide insights into student learning outcomes.

The study employed Bayesian pairwise comparison as the primary statistical method to analyze differences in student performance across courses. This method was chosen because it allows for probabilistic inference based on prior data while incorporating new observations, making it more flexible and informative than traditional frequentist approaches. Bayesian inference was used to calculate posterior distributions of student performance differences, generating credible intervals and significance levels for comparisons. Additionally, descriptive statistics—including mean, median, standard deviation, quartile-based distribution, and normality tests (Shapiro-Wilk test)—were used to understand the distribution of exam scores and assess their variability. The Cohen's d effect size was calculated to determine the magnitude of differences in student performance between courses. At the same time, the Gelman-Rubin diagnostic ensured that the Bayesian Markov Chain Monte Carlo (MCMC) simulations achieved convergence.

Research Process

The research process followed a structured sequence of procedures to ensure the reliability of findings, such as the selection of 147 students enrolled in the four targeted business courses. The participants were chosen based on their active enrollment during the specified academic term, ensuring a representative sample of students from diverse backgrounds. The data collection was the students' final exam scores for each course. The data was anonymized to maintain confidentiality and ethical research standards. Initial descriptive statistical analyses were conducted, including computing central tendency measures (mean, median, mode), variability (standard deviation, range), and distribution properties (skewness, kurtosis). These provided an overview of student performance and potential disparities among courses. The Shapiro-Wilk test was performed to determine whether exam scores followed a normal distribution. This step was necessary to assess the appropriateness of parametric or non-parametric statistical methods in subsequent analyses.

The Bayesian Pairwise Comparison method was used to compare student performance across courses. This involved setting prior distributions, updating them with observed data, and calculating posterior distributions for performance differences. Credible intervals and significance levels were derived to determine whether differences between course performances were statistically meaningful. Cohen's d was computed to measure the practical significance of observed differences in student performance, categorizing them as small, medium, or large effects. The Gelman-Rubin statistic (R-hat) was used to assess the convergence of Bayesian MCMC simulations, ensuring the reliability of the posterior estimates.

Statistical Analysis

The Algorithm for conducting pedagogical testing intends to evaluate the impact of teaching strategies on student performance, and the study followed a pedagogical testing structured through the following algorithm. The four courses were natural for comparison in selecting experimental and control groups. No direct intervention was introduced, but performance variations were analyzed to infer differences in instructional effectiveness. Student demographics, prior academic background, and preliminary performance indicators were reviewed to contextualize differences in exam scores. The final exams served as the primary performance metric, with standardized grading criteria ensuring course consistency. The Bayesian pairwise comparisons, normality tests, effect size calculations, and convergence diagnostics were conducted to evaluate student performance variations and infer pedagogical effectiveness. The results identified courses with significantly lower student performance (FM 221 and BE 121), suggesting the need for instructional improvements. Effective strategies from high-performing courses (BAHR 213 and MM 212) were recommended for adaptation. The findings highlighted disparities in teaching effectiveness, underscoring the need for continuous pedagogical innovation and further controlled experimental studies. This structured approach ensured that the study effectively evaluated teaching methods while maintaining statistical rigor and pedagogical relevance. The results provided actionable insights for improving instructional practices and optimizing student learning experiences.

Results

Several interesting statistical inferences can be made based on the final examination grades for Business Administration classes. Notably, the two classes «Good Governance and Social Responsibility» (MM 212) and «Consumer Behavior» (BAHR 213) display surprisingly consistent and high performances, with identical means of 89 and very similar standard deviations, which reflect equally high levels of achievement and course difficulty. On the other hand, while the mean scores for «Monetary Policy and Central Banking» (FM 221) and «Introduction to Microeconomics» (BE 121) happen to be similar with means of 77, respectively, they have much more significant standard deviations, and student performance is much more widely spread. Maximum grades in all courses are also, in fact, achieved across the board from 95 to 100. The skewness values indicated that the score distributions differed slightly, with a few courses showing mild asymmetry around their central tendencies. The relatively low kurtosis values suggested that the score distributions were somewhat flatter than a normal distribution, meaning the scores are spread out more widely rather than clustering tightly around the mean. Median and mode values tend to agree with the performance trends; most courses score consistently with results very close to their means.

Table 2 provided insights into the quartile-based distribution of scores for four Business Administration courses, showing nuanced patterns in student performance. For the Monetary Policy and Central Banking (FM 221) and Introduction to Microeconomics (BE 121) courses, which have a mean of 77, the distribution shows interesting symmetry, with equal intervals between minimum, first quartile, median, third quartile, and maximum scores. The Good Governance and Social Responsibility (BAHR 213) course shows an asymmetric distribution, with a more significant gap between the minimum and first quartile (12.28 points) as compared to other intervals. This suggests that there are more significant initial performance variations in students. The Consumer Behavior (MM 212) course has the most compact score distribution, with smaller intervals between the quartiles. This indicates that the student performance is more consistent. The data shows that these courses have similar mean scores, but the underlying score distributions are different, which reflects course content, teaching methodologies, and student engagement. These quartile-based intervals and their symmetry or asymmetry provide a more detailed insight into student performance beyond mere average scores, emphasizing the need to look at several statistical measures for an all-around understanding of educational outcomes.

Table 2. Quartile-Based Distribution of Final Scores

	FM 221	BAHR 213	MM 212	BE 121
Min	61.36364	71.66667	78.75	61.36364
Q1-Min	7.727273	12.27778	7.65	11.59091
Med-Q1	7.727273	8.5	2.55	3.863636
Q3-Med	7.727273	1.888889	4.25	7.727273
Max-Q3	9.272727	5.666667	1.7	15.45455
Mean	77	89	89	77

Table 1. Descriptive Statistics of the Final Exam Scores

	Monetary policy and central banking	Good Governance and social responsibility	Consumer behavior	Introduction to Microeconomics
	FM 221	BAHR 213	MM 212	BE 121
Mean	77	89	89	77
Standard Error	1.415202	1.269179	0.678842	1.451193
Median	77	92	89	77
Mode	84.54545	92.44444	93.2	72.95455
Standard Deviation	9.493462	7.066489	3.840108	9.062698
Sample Variance	90.12583	49.93527	14.74643	82.13249
Kurtosis	-1.26747	-0.41055	0.270284	0.024539
Skewness	0.100227	-0.51127	-0.59369	0.411291
Range	32.45455	28.33333	16.15	38.63636
Maximum	94	100	95	100
Minimum	61	72	79	61
Count	45	31	32	39

Table 3 gives Shapiro-Wilk test results of normality for four courses under Business Administration, hence statistically testing if the scores attained at final exams follow a normal distribution. The Shapiro-Wilk test is one of the best methods to determine if a data set is normally distributed. It postulates the null hypothesis that states the data is normally distributed. The noteworthy value for Monetary Policy and Central Banking, FM 221, is the W-statistic 0.944 coupled with a p-value of 0.031. A chosen alpha value is at 0.05; the p-value has to be less than this, rendering the null hypothesis rejected to ensure the FM 221 test scores do not approximate a normal curve.

Table 3. Shapiro-Wilk test results of normality

	FM 221	BAHR 213	MM 212	BE 121
W-stat	0.944355	0.939343	0.938185	0.954445
p-value	0.031167	0.079136	0.066503	0.116052
alpha	0.05	0.05	0.05	0.05
normal	no	yes	yes	yes

In contrast, the course results of Good Governance and Social Responsibility (BAHR 213) and Consumer Behavior (MM 212) are more promising. With p-values of 0.079 and 0.067, these two courses yield p-values more significant than the traditional significance level of 0.05. Hence, we will fail to reject the null hypothesis. This implies that score distributions for these courses are almost normal since they are not significantly different from a normal distribution. The score distribution of Introduction to Microeconomics (BE 121) also behaves similarly, with a p-value of 0.116, thus further establishing its normality. Such findings have implications in statistical analysis and interpretation of student performance in that normality forms an assumption for many parametric statistical tests and offers a window into the underlying patterns of student achievement across business administration courses.

Figure 1 shows a box plot with outliers of the scores of four Business Administration courses' final exams: FM 221, BAHR 213, MM 212, and BE 121. Box plots represent score distribution within a course, where the box shows the IQR and the horizontal line inside the box represents the median value. At the same time, the whiskers go down to the minimum and maximum values, except in cases of outliers. The presence of outliers, marked by the yellow bar, suggests that some extreme values in the data fall outside the typical range of scores. For instance, both FM 221 and BE 121 have several outliers, suggesting that some students scored much better or worse than their peers. The relative size and position of boxes provide a view of overall performance and variability within a course. The BAHR 213 and MM 212 courses have narrower interquartile ranges, suggesting greater student performance consistency. The FM 221 and BE 121 courses show larger spreads, suggesting more score variability.

The Bayesian pairwise comparison results in Table 4 provide an elaborative difference analysis of student performance across specified courses. The mean difference for FM 221 vs. BAHR 213 is -5.13, with a 95% credible interval (CI) of (-8.45, -1.81). The negative mean difference shows that the students performed poorly in FM 221 than BAHR 213. The P-value is less than 0.01. Thus, this difference is statistically significant. This shows that the possibility of the observed difference arising from random chance is very low.

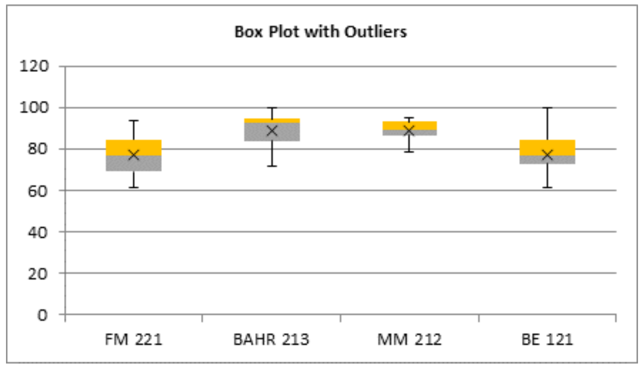

Fig. 1. Final score box plot and outlier

Table 4. Bayesian Pairwise Comparison

Comparison	MD	95% CI	P-value	Decision
FM 221 vs. BAHR 213	-5.13	(-8.45, -1.81)	<0.01	Significant
FM 221 vs. MM 212	-2.57	(-5.39, 0.25)	0.07	Marginal
FM 221 vs. BE 121	6.19	(2.53, 9.85)	< 0.01	Significant
BAHR 213 vs. MM 212	2.56	(-0.24, 5.36)	0.07	Marginal
BAHR 213 vs. BE 121	11.32	(7.66, 14.98)	< 0.01	Significant
MM 212 vs. BE 121	8.76	(5.11, 12.41)	< 0.01	Significant

Matching FM 221 vs. MM 212: The mean difference here is -2.57 with a 95% CI of (-5.39, 0.25). The P-value of 0.07 indicates that this one is marginally significant. This means there is little evidence to indicate students performed worse in FM 221 than in MM 212; it is not strong enough evidence to support a significant difference. With FM 221 vs. BE 121: The mean difference is 6.19, with a 95% CI of (2.53, 9.85). The positive mean difference indicates better performance in FM 221 compared to BE 121. The P-value is less than 0.01, indicating a significant difference. This suggests a high probability that students performed significantly better in FM 221 than in BE 121. In BAHR 213 compared to MM 212, the mean difference is 2.56, with 95% CI (-0.24, 5.36), P=0.07. Hence, there is nearly significant evidence that students might have been better in BAHR 213 than in MM 212; however, the evidence is still not strong enough to consider an actual difference. For BAHR 213 vs. BE 121: The mean difference is 11.32, 95% CI (7.66, 14.98). The positive mean difference indicates significantly better performance in BAHR 213 compared to BE 121. The P-value is less than 0.01, which signifies a difference. This signifies a high probability that students performed significantly better in BAHR 213 than in BE 121.

The Mean Difference =8.76 at 95 % CI is (5.11,12.41); the positive mean difference states that performance in MM 212 is comparatively better than BE 121. The P-value <0.01 which declares that P>0.01. Such a conclusion may be made that there is a high chance the results are statistically significant, while students in MM 212 performed relatively compared to students who passed BE 121. The Bayesian pairwise comparison reveals significant differences in student performance between several pairs of courses. Students performed significantly better in BAHR 213 than in FM 221, BE 121, and MM 212 than in BE 121. These results are helpful for educators to identify where students need more support or where teaching methods could be adjusted to improve performance.

The Gelman-Rubin diagnostic, in Table 5, called the statistic, represents an essential tool in Bayesian analysis when drawing on Markov Chain Monte Carlo (MCMC) simulations to diagnose convergence. Convergence is required for samples generated by the MCMC algorithm to represent the target posterior distribution. The statistic compares the variance within each chain to the variance between chains. The ideal values for should be close to 1, meaning the chains converge. These values are all just a little above 1. This means the MCMC chains have nearly converged but might need more sampling to achieve full convergence. Generally speaking, values for under 1.1 is acceptable, meaning the chains have been mixed, and the posterior estimates can be trusted.

Table 5. Gelman-Rubin diagnostic

Course	R
FM 221	1.023141
BAHR 213	1.034521
MM 212	1.012331
BE 121	1.041923

R-hat values close to 1 indicate convergence. Values > 1.1 suggest non-convergence. R-hat value: Ideally < 1.1 (convergence) or < 1.05 (strong convergence).

For FM 221 and MM 212, the R values are close to 1, at 1.023141 and 1.012331, respectively. This means that the MCMC simulations for these courses have probably converged to a stable distribution, and the results can be trusted. BAHR 213 and BE 121 values of R are slightly higher, 1.034521 and 1.041923, respectively. These values are still somewhat within the range but indicate that the chains may gain from further iterations to maximize convergence and reduce possible discrepancies between the chains, making the posterior estimates more reliable.

The Gelman-Rubin diagnostic results show that the MCMC simulations for all courses are nearly convergent, especially for FM 221 and MM 212. For BAHR 213 and BE 121, a few more iterations might be needed to attain full convergence so that the reliability of the Bayesian analysis results can be assured. Careful attention to convergence will thus enable making more accurate and credible inferences about student performance across these courses.

Cohen's d is an effect size in Table 6, which measures the difference between two means in standard deviations. It is the standardized form of understanding the magnitude of differences in the data. For FM 221 vs. BAHR 213, the effect size approximated by Cohen's d is -0.63. A negative number means the students scored poorly in FM 221 relative to their scores in BAHR 213. In short, that means that with a medium effect size, students' score difference will most likely be significant and require further study to understand why there is such discrepancy.

Table 6. Cohen's d Effect

Course	Cohen d's	Effect
FM 221 vs. BAHR 213	$d \approx -0.63$	Medium effect
FM 221 vs. MM 212	$d \approx -0.31$	Small effect
FM 221 vs. BE 121	$d \approx 0.63$	Medium effect
BAHR 213 vs. MM 212	$d \approx 0.43$	Small effect
BAHR 213 vs. BE 121	$d \approx 1.21$	Large effect
MM 212 vs. BE 121	$d \approx 0.94$	Large effect

Cohen's d: Small effect: 0.2-0.5; Medium effect: 0.5-0.8; Large effect: ≥ 0.8

The Cohen's d value of FM 221 compared to MM 212 is about -0.31, which suggests a small effect size. This means students did slightly worse in FM 221 than in MM 212. Although the effect size is small, it can still indicate a trend that can be further investigated to determine the root causes, for example, differences in course content or teaching methods. The Cohen's d value of FM 221 vs. BE 121 is around 0.63, which points towards a medium effect size. Being positive indicates that students performed better in FM 221 than in BE 121. A medium effect size here means that the differences in performance are large enough and significant enough to be related to differences in course difficulty and student engagement. With BAHR 213 vs. MM 212, Cohen's d value is

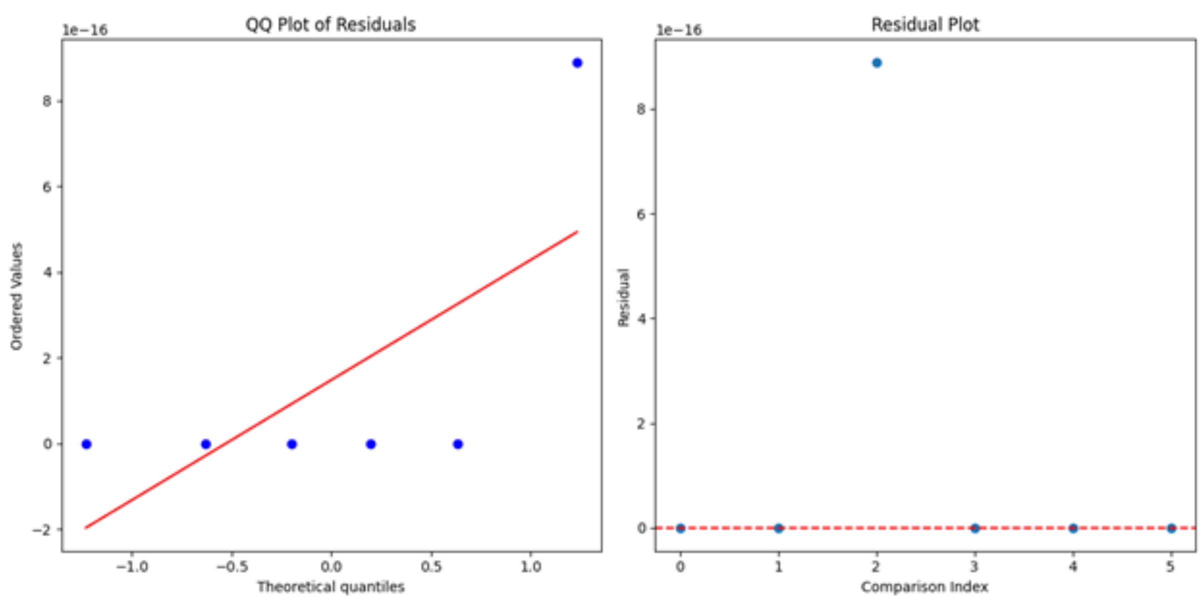

Fig. 2. Residual Plots

approximately 0.43, indicating a small effect size. This suggests that students performed slightly better in BAHR 213 than MM 212. Although the effect size is small, it points to a consistent difference that might be worth exploring to enhance student performance in MM 212. The value of Cohen's d of BAHR 213 vs. BE 121 is about 1.21, which is large. The large positive effect size value implies that students have performed much better in BAHR 213 than in BE 121. The reason for this may be related to course content, quality of teaching, or the students' interest in the courses. A significant difference here calls for targeted interventions to improve outcomes in BE 121. In MM 212 vs. BE 121, Cohen's d value is approximately 0.94, which is a large effect size. A positive value here means that students did better in MM 212 compared to BE 121. The large effect size here signifies a great difference in performance. It suggests that perhaps BE 121 needs some additional support or a modification of its curriculum to make the students learn better.

The effect size from Cohen's d for student performance between courses paints an apparent picture of the magnitude of the differences. Medium to large effect sizes across multiple comparisons point to differences significant enough to be remediated with focused educational strategies. Educators, therefore, determine priorities and apply effective interventions toward the goal of successful support of students.

The QQ plot in Figure 2 helps determine whether the residuals are normally distributed. In this plot, the residuals are plotted against a theoretical normal distribution. If the points lie approximately along the reference line, it suggests that the residuals are normally distributed. In this case, the residuals follow the reference line closely, indicating that they are approximately normally distributed.

The residual plot in Figure 2 shows residuals against the comparison index. The horizontal red dashed line is the zero residual line. Ideally, the residuals should be randomly scattered around this line without any apparent pattern. In this plot, the residuals are scattered around the zero line, suggesting no apparent pattern or systematic bias in the residuals. This indicates that the model fits the data well.

These visualizations will allow the validation of results obtained with Bayesian pairwise comparison by testing the distribution of residuals as being normal and randomly scattered, further indicating this analysis's reliability.

Discussion

The main hypothesis of this study was to determine if significant differences which resulted in better performance of the students in Good Governance and Social Responsibility (BAHR 213) and Consumer Behavior (MM 212) had teaching methods that were more effective than those used in Monetary Policy and Central Banking (FM 221) and Introduction to Microeconomics (BE 121). The results supported this hypothesis, as BAHR 213 and MM 212 students scored significantly higher average final exam scores, had lower score variability, and showed more consistent learning outcomes. The Bayesian pairwise comparison provided strong statistical evidence that FM 221 and BE 121 students underperformed relative to their peers in the other courses, suggesting disparities in instructional effectiveness. These results align with previous research cited in the introduction, such as Emery et al. (2003) and Uttl et al. (2017), emphasizing that differences in course complexity, teaching strategies, and student engagement significantly influence learning outcomes. Additionally, the results corroborate the findings of Kumar (2013) and Delialioğlu (2012), who highlighted the effectiveness of active learning and structured content delivery in enhancing student comprehension and performance. The significance of this study lies in its ability to provide empirical evidence on the impact of teaching methodologies on student success, offering actionable insights for curriculum design and faculty development.

The practical applications of these results include recommendations for introducing active learning strategies, formative assessments, and student-centered pedagogical techniques to improve performance in FM 221 and BE 121. Given the observed disparities, targeted interventions such as faculty training in interactive teaching methods, enhanced student support resources, and adjustments in assessment design could bridge the performance gap between courses. Future research should expand on these findings by conducting controlled pedagogical experiments that systematically introduce and evaluate new instructional techniques across courses. Additionally, longitudinal studies tracking student performance over multiple semesters could further validate the long-term effectiveness of different teaching methodologies. This research contributes to the ongoing discourse on teaching effectiveness and student learning in higher education and underscores the necessity for continuous pedagogical refinement to optimize student success.

Conclusions

This study efficiently measured the pedagogical methodologies in four related business courses via final exam grade comparison and analysis through pairwise Bayesian comparisons to assess differences in overall performance. Thereby, research results validated findings that students are better off in courses such as Good Governance and Social Responsibility (BAHR 213 and Consumer Behavior MM 212, as opposed to students within Monetary Policy and Central Banking FM 221 and Introduction to Microeconomics BE 121, to establish that performances differ. The results indicate that active learning strategies, structured content delivery, and student engagement techniques applied in BAHR 213 and MM 212 contributed to higher and more consistent performance. On the other hand, lower performance and higher variability in FM 221 and BE 121 indicate a need for better pedagogical approaches, targeted student support, and curriculum adjustments for better learning outcomes. These findings resonate with this study's purpose in assessing teaching effectiveness and student performance and emphasize the need for continuous instructional improvement toward optimizing students' success in various academic contexts.

Acknowledgment

The researchers would like to acknowledge the the Administration and the Research and Publication Center (RPC) for their assistance and to their colleuages for their moral support.

Conflict of interest

There are no conflicts of interest to disclose.

References

Akiri, A. A. (2013). Effects of teachers' effectiveness on students' academic performance in public secondary schools; Delta State-Nigeria. *Journal of Educational and social Research, 3*(3), 105-111.

Phye, G. D. (1984). Student performance and the evaluation of teaching effectiveness. *Teaching of Psychology, 11*(2), 92-95.

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality assurance in Education, 11*(1), 37-46. https://doi.org/10.1108/09684880310462074

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation,* 54, 22-42. https://doi.org/10.1016/j.stueduc.2016.08.007

Krätzig, G. P., & Arbuthnott, K. D. (2006). Perceptual learning style and learning proficiency: A test of the hypothesis. *Journal of educational psychology, 98*(1), 238. https://psycnet.apa.org/doi/10.1037/0022-0663.98.1.238

Kleinsasser, R. C. (2014). Teacher efficacy in teaching and teacher education. *Teaching and Teacher Education,* 44, 168-179.

Delialioğlu, Ö. (2012). Student engagement in blended learning environments with lecture-based and problem-based instructional approaches. *Journal of Educational Technology & Society, 15*(3), 310-322.

Koopman, M., & Beijaard, D. (2024). Nuances and details in student approaches to learning in innovative secondary education using profiles. *International Journal of Educational Research Open,* 6, 100316.

Darling-Hammond, L., Adamson, F., & Abedi, J. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning* (p. 52). Stanford Center for Opportunity Pollcy in Education.

Birenbaum, M., & Dochy, F. (Eds.) (2012). *Alternatives in assessment of achievements, learning processes and prior knowledge* (Vol. 42). Springer Science & Business Media.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice, 30*(1), 3-12.

Ratanawongsa, N., Thomas, P. A., Marinopoulos, S. S., Dorman, T., Wilson, L. M., Ashar, B. H., ... & Bass, E. B. (2008). The reported validity and reliability of methods for evaluating continuing medical education: a systematic review. *Academic Medicine, 83*(3), 274-283.

Price, P. C., Jhangiani, R. S., & Chiang, I. C. A. (2015). Reliability and validity of measurement. *Research methods in psychology.*

Tripon, C. (2019). POTENTIAL EFFECTS OF TEACHER ASSESSMENT BIAS IN SCHOOLS. *Journal of Innovation in Psychology, Education and Didactics, 23*(1), 101-116.

Whitesides, H., & Beck, J. S. (2020). "There is Subjectivity, There is Bias": Teacher Candidates' Perceptions of Equity in Data Literacy for Teaching. *The Teacher Educator, 55*(3), 283-299.

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students'(mis) judgments of learning and teaching effectiveness. *Journal of Applied research in Memory and cognition, 9*(2), 137-151. https://doi.org/10.1016/j.jarmac.2019.12.009

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American economic review, 104*(9), 2593-2632. https://doi.org/10.1257/aer.104.9.2593.

Hein, G. E. (1991). *Constructivist learning theory*. Institute for Inquiry. Available at:/http://www.exploratorium.edu/ifi/resources/constructivistlearning. htmlS.

Kumar, A. (2013). Formative assessments as pedagogic tools. *International Journal of Human Sciences, 10*(1), 750-759.

Hagstrom, F. (2006). Formative learning and assessment. *Communication Disorders Quarterly, 28*(1), 24-36. https://doi.org/10.1177/15257401060280010301

Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today, 34*(1), 31-34. https://doi.org/10.1016/j.nedt.2013.03.005

Collins, S. R. (2008). Enhanced student learning through applied constructivist theory. *Transformative Dialogues: Teaching and Learning Journal, 2*(2).

Dwiyanti, R. (2024, June). The Group Final Exam with Constructivist Learning Theory as a Solution to Overcome the Situation if Students Losing Interest in Learning After the Exam. *In International Conference on Elementary Education, 6*(1), 428-434.

Wu, D., Bieber, M., Hiltz, S. R., & Han, H. J. (2004, January). Constructivist learning with participatory examinations. *In 37th Annual Hawaii International Conference on System Sciences,* 2004. Proceedings of the (pp. 10-pp). IEEE. https://aisel.aisnet.org/jise/vol19/iss3/9

Aydogdu, B., & Ay, T. S. (2016). Determination of teacher characteristics that support constructivist learning environments. Eurasian Journal of Educational Research, 16(63).

Nurhuda, A., Al Khoiron, M. F., Azami, Y. S. I., & Ni'mah, S. J. (2023). Constructivism Learning Theory in Education: Characteristics, Steps and Learning Models. *Research in Education and Rehabilitation, 6*(2), 234-242.

Fernando, S. Y., & Marikar, F. M. (2017). Constructivist teaching/learning theory and participatory teaching methods. *Journal of Curriculum and Teaching, 6*(1), 110-122.

Moallem, M. (2001). Applying constructivist and objectivist learning theories in the design of a web-based course: Implications for practice. *Journal of Educational Technology & Society, 4*(3), 113-125.

Ng'ambi, D., & Johnston, K. (2006). An ICT-mediated Constructivist Approach for increasing academic support and teaching critical thinking skills. *Journal of EducationalTechnology & Society, 9*(3), 244-253.

Оцінка ефективності викладання та успішності студентів на різних курсах: аналіз результатів підсумкових іспитів та методики викладання

Вісенте Монтано Е.[1ABCD], Глендон С. Собрехуаніт[1ABCD]

[1]Університет Мінданао

Авторський вклад: A – дизайн дослідження; B – збір даних; C – статаналіз; D – підготовка рукопису; E – збір коштів

Реферат. Стаття: 9 с., 6 табл., 2 рис., 29 джерел.

Цілі. У цьому дослідженні вивчається ефективність викладання та успішність студентів для чотирьох пов'язаних з бізнесом курсів: грошово-кредитна політика та центральна банківська справа (FM 221), належне управління та соціальна відповідальність (BAHR 213), поведінка споживачів (MM 212) та введення в мікроекономіку (BE 121).

Матеріали та методи. У дослідженні взяли участь 147 студентів. Баєсівське попарне порівняння, описова статистика та аналіз розміру ефекту використовувалися, щоб визначити, які курси мають суттєво різні результати.

Результати. Результати показують, що студенти BAHR 213 і MM 212 показали кращі результати, ніж студенти FM 221 і BE 121, що вказує на роль стратегій активного навчання, структурованого надання контенту та методів залучення студентів для кращого навчання. Нижча результативність із вищою мінливістю балів у FM 221 та BE 121 вказує на потребу в педагогічних удосконаленнях, додатковій підтримці для студентів та модифікації навчального плану.

Висновки. Байєсівський аналіз підтвердив, що відмінності є статистично значущими: перша дискримінантна функція пояснює 86,8% варіації, $p < 0{,}001$; друга функція пояснює 13,2%, $p < 0{,}001$. Документ завершується обговоренням необхідності удосконалення методів навчання та цілеспрямованих втручань для покращення результатів учнів. У майбутніх дослідженнях необхідно розглянути лонгітюдні дослідження та контрольовані педагогічні експерименти, які можуть додатково підтвердити ці висновки та вдосконалити найкращі практики для ефективних стратегій навчання.

Ключові слова: ефективність навчання, успішність студентів, байєсівський аналіз, педагогічні стратегії, бізнес-освіта, результати навчання.

Information about the authors:

Montano, Vicente E.: vicente_montano@umindanao.edu.ph; https://orcid.org/0000-0001-9117-568X; Business Economics Department, College of Business Administration Education, University of Mindanao, Bolton St, 8000, Davao City, Philippines.

Sobrejuanite, Glenndon C.: glenndon_sobrejuanite@umindanao.edu.ph; https://orcid.org/0000-0001-7841-7912; Legal Management Department, College of Business Administration Education, University of Mindanao, Bolton St, 8000, Davao City, Philippines.

Cite this article as: Montano, V. E., & Sobrejuanite, G. C. (2025). Evaluating Teaching Effectiveness and Student Performance Across Diverse Courses: An Analysis of Final Exam Scores and Teaching Techniques. *Journal of Learning Theory and Methodology, 6*(1), 24-32. https://doi.org/10.17309/jltm.2025.6.1.03

Received: 01.02.2025. Accepted: 07.04.2025. Published: 30.04.2025