Revista Española
de Educación Médica

UNIVERSIDAD DE MURCIA

Academic Performance and Student Satisfaction under Two Teaching Approaches in Neuroanatomy among Medical Students: A Quasi-Experimental Difference-in-Differences Study.

Rendimiento académico y satisfacción estudiantil bajo dos enfoques docentes en neuroanatomía entre estudiantes de Medicina: un estudio cuasiexperimental de diferencias en diferencias.

Juan Francisco García Granado[1]*, Ayoze Nauzet González Hernández[2].

[1] Hospital Universitario de Gran Canaria Doctor Negrín, España, juuanfra21@gmail.com, https://orcid.org/0000-0003-0670-7069

[2] Hospital Universitario de Gran Canaria Doctor Negrín, Universidad Fernando Pessoa Canarias, España, ayozegonzalez@hotmail.com, https://orcid.org/0000-0003-2986-3603.

* Correspondencia: juuanfra21@gmail.com

Recibido: 16/5/26; Aceptado: 3/6/26; Publicado: 5/6/26

Abstract.
Objective. This study aimed to evaluate differences associated with two teaching methodologies on academic performance and student evaluation of teaching quality in neuroanatomy, using physiology as a control group. **Methodology.** A quasi-experimental study compared two academic years (2023–2024 and 2024–2025). Neuroanatomy changed from a flipped classroom to a traditional lecture-based teaching format, while physiology remained traditional lecture-based teaching in both years. Academic performance was assessed using final grades (primary outcome), examination scores, and continuous assessment. Student evaluation of teaching quality was measured using institutional surveys and normalized to a 0–100 scale. Comparisons were performed using the Mann–Whitney U test. Difference-in-differences (DiD) analyses were conducted using linear regression models including academic year, subject, and their interaction. **Results.** A total of 87 neuroanatomy students and 86 physiology students were included. In neuroanatomy, final grades decreased from 4.44 ± 2.33 to 3.02 ± 2.21 ($p = 0.006$), with significant reductions in examination scores ($p = 0.007$) and continuous assessment ($p < 0.001$). Similar declines were observed in physiology (final grades: 6.20 ± 1.27 vs. 3.39 ± 2.98; $p < 0.001$). In the DiD analysis, the interaction term showed a non-significant effect ($\beta = -1.387$; 95% CI [-2.814, 0.039]; $p = 0.057$), suggesting that no statistically significant independent effect of teaching methodology on performance could be demonstrated after adjustment for cohort effects. In contrast, student satisfaction was higher in 2024–2025 after normalization, with DiD analyses indicating improved engagement ($\beta = -7.17$; $p < 0.05$) and a trend toward higher perceived clarity. **Conclusions.** No teaching methodology demonstrated statistically significant superiority in academic performance after adjustment for cohort effects. However, the methodology implemented in 2024–2025 was associated with higher student satisfaction, particularly in engagement. These findings suggest a dissociation between perceived and objective learning outcomes and highlight the importance of controlled designs in educational research.

Keywords: Flipped classroom, Traditional lecture-based teaching, Neuroanatomy, Difference-in-Differences, Student satisfaction.

Resumen.
Objetivo. Este estudio tuvo como objetivo evaluar las diferencias asociadas a dos metodologías docentes sobre el rendimiento académico y la evaluación de la calidad docente por parte de los estudiantes en neuroanatomía, utilizando fisiología como grupo control. **Métodos.** Se realizó un estudio cuasiexperimental comparando dos cursos académicos (2023–2024 y 2024–2025). La asignatura de neuroanatomía cambió de una metodología de aula invertida a un formato tradicional basado en clases magistrales, mientras que fisiología mantuvo una metodología convencional en ambos cursos. El rendimiento académico se evaluó mediante las calificaciones finales (resultado principal), las puntuaciones de los exámenes y la evaluación continua. La valoración estudiantil de la calidad docente se midió mediante encuestas institucionales y se normalizó a una escala de 0–100. Las comparaciones se realizaron mediante la prueba U de Mann–Whitney. Los análisis de diferencias en diferencias (Difference-in-Differences, DiD) se llevaron a cabo utilizando modelos de regresión lineal que incluían el curso académico, la asignatura y su interacción. **Resultados.** Se incluyeron un total de 87 estudiantes de neuroanatomía y 86 de fisiología. En neuroanatomía, las calificaciones finales disminuyeron de 4.44 ± 2.33 a 3.02 ± 2.21 ($p = 0.006$), con reducciones significativas tanto en las puntuaciones de los exámenes ($p = 0.007$) como en la evaluación continua ($p < 0.001$). Descensos similares se observaron en fisiología (calificaciones finales: 6.20 ± 1.27 frente a 3.39 ± 2.98; $p < 0.001$). En el análisis DiD, el término de interacción mostró un efecto no significativo ($\beta = -1.387$; IC 95% [-2.814, 0.039]; $p = 0.057$), lo que sugiere la ausencia de un impacto independiente de la metodología docente sobre el rendimiento académico. En contraste, la satisfacción estudiantil fue mayor en 2024–2025 tras la normalización de las puntuaciones, mostrando los análisis DiD una mejor implicación estudiantil ($\beta = -7.17$; $p < 0.05$) y una tendencia hacia una mayor claridad percibida. **Conclusiones.** Ninguna metodología docente demostró superioridad en el rendimiento académico tras ajustar por efectos de cohorte. Sin embargo, la metodología implementada en 2024–2025 se asoció con una mayor satisfacción estudiantil, especialmente en la implicación del alumnado. Estos hallazgos sugieren una disociación entre los resultados de aprendizaje percibidos y los objetivos, y resaltan la importancia de utilizar diseños controlados en investigación educativa.

Palabras clave: Aula invertida, Enseñanza tradicional, Neuroanatomía, Diferencias en diferencias, Satisfacción estudiantil.

1. Introduction

The rapid evolution of educational strategies in higher education has led to a growing interest in active learning approaches aimed at enhancing student engagement and improving learning outcomes. In medical education, Traditional lecture-based teaching has progressively been complemented or replaced by more student-centered methodologies designed to promote deeper understanding and critical thinking (1-2). Among these approaches, the flipped classroom model has gained increasing attention. This model shifts the acquisition of theoretical knowledge outside the classroom, allowing in-person sessions to be used for interactive activities such as problem-solving, discussion, and application of concepts. By promoting active participation, the flipped classroom has been proposed as an effective strategy to improve both knowledge retention and higher-order cognitive skills (3-4).

However, the evidence regarding the effectiveness of flipped classroom methodologies in medical education remains heterogeneous. While several studies have reported improvements in academic performance and student satisfaction compared to traditional lecture-based teaching approaches (5), others have shown no significant differences or have highlighted challenges related to student adaptation and increased workload (6). Furthermore, student evaluation of teaching quality, an important component of educational assessment, may also be influenced by teaching methodology, although this relationship is complex and potentially confounded by factors such as academic performance and cohort characteristics (7-8). Importantly, objective educational outcomes and subjective student perceptions do not necessarily evolve in parallel. A teaching methodology may be associated with higher levels of satisfaction, engagement, or perceived teaching quality without producing measurable improvements in academic performance. Conversely, educational approaches that promote deeper learning and knowledge acquisition may not always be perceived more favorably by students, particularly when they require greater effort, autonomy, or adaptation. Therefore, evaluating both academic outcomes and student perceptions is essential to obtain a comprehensive understanding of the educational impact of different teaching strategies.

Neuroanatomy represents a particularly demanding subject within medical curricula due to its high level of abstraction and reliance on spatial reasoning. These characteristics make it especially sensitive to teaching strategies, and therefore an ideal context in which to evaluate the impact of innovative educational approaches.

A key methodological limitation in many studies assessing educational interventions is the lack of adequate control for inter-cohort variability. Comparisons across different academic years may be biased by differences in student characteristics, baseline academic ability, or contextual factors unrelated to the intervention (9). To address this issue, quasi-experimental designs incorporating control groups, such as parallel courses with stable teaching methodologies, have been recommended as a robust approach to isolate the true effect of educational changes (10).

Among these approaches, Difference-in-Differences (DiD) analysis allows researchers to estimate the effect of an intervention by comparing changes over time in an intervention group with simultaneous changes observed in a control group. In the present study, physiology served as a concurrent control subject because its teaching methodology remained unchanged across academic years. By focusing on differences in trends rather than simple between-group comparisons, the DiD approach can partially account for cohort-related variability and secular changes affecting both groups. However, a key assumption of this method is that, in the absence of the intervention, both groups would have followed comparable underlying trends over time. As with many educational studies, this assumption cannot be formally verified because only two academic years were available, and therefore the findings should be interpreted with appropriate caution.

In this context, the present study aims to evaluate differences associated with two teaching approaches under a quasi-experimental framework on academic performance and student evaluation of teaching quality in a neuroanatomy course. To minimize cohort-related bias, a parallel course (physiology) with a stable teaching methodology across the same academic years was included as a control group, allowing for a controlled quasi-experimental comparison.

2. Methods

2.1. Study design and setting

A quasi-experimental study with a difference-in-differences (DiD) approach was conducted to evaluate the impact of a change in teaching modality on academic performance and student satisfaction in undergraduate medical education. The study compared two consecutive academic years (2023–2024 and 2024–2025) in a neuroanatomy course in which a change in teaching methodology occurred as part of routine educational practice. Specifically, the course was delivered using a flipped classroom model during the 2023–2024 academic year and a traditional lecture-based teaching format during the 2024–2025 academic year. The present investigation was conducted retrospectively using anonymized academic and institutional evaluation data and was not originally planned as a prospective educational intervention study. In contrast, the physiology course was taught using a traditional lecture-based teaching methodology in both academic years and was included as a control group.

This design allowed for a controlled before–after comparison using a difference-in-differences (DiD) approach to control for inter-cohort variability. The Physiology course was selected as a parallel control group because both subjects (Neuroanatomy and Physiology) are taught concurrently to the same cohort of medical students within the same academic year. This minimizes the influence of external factors (such as differences in student characteristics, motivation, or academic context) that could affect both subjects equally.

The study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines, as it constitutes a retrospective quasi-experimental observational study. Data were obtained from institutional academic records and routine student teaching evaluation surveys. Academic records included final grades and their individual assessment components, whereas teaching evaluation surveys provided information on students' perceptions of teaching quality and satisfaction. All data had originally been collected as part of routine educational activities and were subsequently extracted for research purposes after completion of the corresponding academic years. Prior to analysis, all datasets were anonymized by removing personal identifiers, and only aggregated academic and survey information was used. In accordance with STROBE recommendations, participant inclusion, data availability, outcome assessment, statistical analyses, and study limitations are explicitly reported to enhance transparency and reproducibility.

2.2. Participants

All students enrolled in the neuroanatomy and physiology courses during the academic years 2023–2024 and 2024–2025 were eligible for inclusion. No sampling was performed, and all available data were considered for analysis. No formal exclusion criteria were applied. However, for specific academic performance analyses, only students with available final grade records were included. Consequently, sample sizes varied slightly across outcomes according to data availability, as detailed in **Figure 1** and the corresponding tables. As the study was based on aggregated academic and evaluation data, individual informed consent was not required.

2.3. Teaching modalities

In the 2023–2024 academic year, neuroanatomy was delivered using a flipped classroom model, in which students reviewed theoretical content prior to classroom sessions through educational materials provided by the teaching staff, including lecture notes, recommended readings, and resources available on the institutional virtual learning platform. Classroom sessions were designed to promote active learning and were dedicated to activities such as guided discussion, problem-solving exercises, application of neuroanatomical concepts, and interactive question-and-answer sessions. Students were expected to prepare the assigned material before attending class and actively participate in the learning activities.

In the 2024–2025 academic year, the course followed a traditional lecture-based teaching model, in which theoretical content was primarily delivered during in-person sessions, complemented by practical and seminar activities.

In contrast, the physiology course maintained a consistent traditional lecture-based teaching methodology across both academic years and was used as a control group to account for potential differences between student cohorts.

2.4. Variables and outcomes

Academic performance was defined as the primary outcome and was assessed using the overall final grade on a 0–10 scale, reflecting the weighted combination of all evaluation components. The final grade was calculated as a composite score, with the final examination accounting for 75% of the total grade, individual coursework for 10%, and group-based activities for 15%. The evaluation system included a final examination, individual coursework consisting of formative assessments, and group-based practical and seminar activities, allowing assessment of both theoretical knowledge and applied learning. The evaluation structure and weighting of assessment components remained unchanged across the two academic years within each subject. Therefore, observed differences in academic performance were not attributable to modifications in grading criteria or assessment weighting.

Secondary outcomes included the examination score as an independent measure of theoretical knowledge, performance in coursework activities, and the proportion of students who passed the course.

Student satisfaction was also included as a secondary outcome and was assessed using institutional teaching evaluation surveys based on Likert-type scales. Due to differences in survey design and scoring systems between academic years (e.g., 1–5 and 1–7 scales), all scores were normalized to a 0–100 scale to allow comparability. Normalization was performed by expressing each observed score as a percentage of its corresponding scale range and was performed using the following formula: Normalized score = ((Observed score − Minimum possible score) / (Maximum possible score − Minimum possible score)) × 100. Survey items were harmonized across academic years by grouping conceptually equivalent questions into predefined domains, including clarity and structure, methodology and teaching quality, engagement and participation, support and accessibility, motivation and inspiration, relationship and professionalism, and organization (table 1). The number of items contributing to each domain varied according to the structure of the institutional surveys in each academic year. Although domains were constructed using conceptually similar items, complete equivalence between questionnaires could not be guaranteed because of differences in item wording, scale structure, and domain composition. Therefore, comparisons of student satisfaction across academic years should be interpreted with caution. Each survey item was assigned to a single domain to avoid duplication and ensure conceptual independence. The unit of analysis was the survey item, and aggregated mean scores were used for all analyses. Domain-level comparisons between academic years were performed by analyzing the distribution of item-level mean scores within each domain.

2.5. Statistical analysis

Descriptive statistics were used to summarize student characteristics and outcomes across courses and academic years. Normality of continuous variables was assessed using the Shapiro–Wilk test together with visual inspection of histograms and Q–Q plots. Continuous variables were expressed as mean and standard deviation or median and interquartile range, depending on their distribution, while categorical variables were presented as frequencies and percentages. Between-group comparisons were performed using Student's t-test or Mann–Whitney U test for continuous variables and chi-square tests for categorical variables. Differences in sample size across analyses were due to missing academic records, unavailable survey responses, or incomplete assessment components. The number of observations included in each analysis is reported explicitly in the corresponding tables.

Data cleaning procedures were applied prior to analysis, and missing or non-applicable values were handled according to institutional grading policies. Special attention was paid to students who were enrolled in the course during both academic periods (2023–2024 and 2024–2025). These cases were identified and assessed to ensure that prior exposure did not bias the comparison of academic performance between teaching methodologies. To estimate the effect of the teaching methodology while controlling for inter-cohort variability, a difference-in-differences (DiD) approach was applied. Regression models included course (neuroanatomy versus physiology), academic year (2023–2024 versus 2024–2025), and their interaction term, which represented the effect of the intervention. Linear regression models were used for continuous outcomes such as overall final grade, examination scores, and normalized satisfaction scores. Logistic regression models were applied for binary outcomes such as pass rates. For satisfaction analyses, DiD models were also applied to the normalized domain-specific scores to assess differences between methodologies while adjusting for cohort effects. Results were reported as beta coefficients or odds ratios with 95% confidence intervals. A two-sided p-value of less than 0.05 was considered statistically significant. All statistical analyses were performed using SPSS version 26 (IBM Corp., Armonk, NY, USA).

Table 1. Classification of student satisfaction survey items into conceptual domains.

Domain	Description	Representative Items
Clarity and Structure	Clarity of explanations, definition of objectives, and overall organization of teaching	"The teacher clearly explains the objectives at the beginning of the course"; "Classes are well prepared"; "Explanations are clear"; "The teacher helps me achieve the course objectives"
Methodology and Teaching Quality	Perceived usefulness of content, teaching strategies, and instructional design	"The teacher shows the usefulness of the content"; "The teacher is dynamic and uses effective teaching methods"; "Learning activities help achieve objectives"; "The teacher provides meaningful content"
Engagement and Participation	Student involvement, active participation, and interest during sessions	"The teacher encourages participation"; "The teacher stimulates interest and participation"; "The teacher generates interest in the subject"
Support and Accessibility	Availability of the instructor and responsiveness to student needs	"The teacher is available to answer questions during class and tutorials"
Motivation and Inspiration	Ability of the teacher to motivate, inspire, and promote personal/professional growth	"The teacher motivates and inspires me"; "The teacher stimulates curiosity"; "The teacher encourages professional development"
Relationship and Professionalism	Teacher–student relationship and interpersonal climate	"The teacher fosters a respectful relationship"; "The teacher treats students appropriately"
Organization	Logistical aspects of the course and use of teaching resources	"The teacher starts classes on time"; "Recommended materials are useful"

2.6. Ethical Considerations

This study was conducted in accordance with institutional policies governing the use of anonymized academic data for research purposes and complied with the principles of the Declaration of Helsinki. All data were obtained from routine institutional academic records and teaching evaluation surveys collected as part of standard educational activities. The study was conducted retrospectively using fully anonymized datasets, and no personal identifiers were available to the investigators at any stage of the analysis. Because the research involved the secondary use of anonymized educational data, did not require direct participant involvement, and did not modify routine educational practice, individual informed consent was not required. Formal ethics committee approval was also not required according to institutional regulations applicable to retrospective studies based exclusively on anonymized academic performance and teaching evaluation data. Confidentiality and data protection were strictly maintained throughout the study in accordance with applicable data protection legislation.

3. Results

3.1. Cohort Description and Academic Performance Overview

A total of 87 students enrolled in the neuroanatomy course were included in the analysis, comprising 40 students in the 2023–2024 academic year and 47 students in the 2024–2025 academic year. Among these, final grade data were available for 84 students (39 in 2023–2024 and 45 in 2024–2025) (figure 1). The small discrepancy between total enrollment and available final grade records reflects the absence of complete academic outcome data for a limited number of students in the institutional database. The mean final grade was 4.44 ± 2.33 in the 2023–2024 cohort and 3.02 ± 2.21 in the 2024–2025 cohort. The median grades were 3.30 and 2.30. This difference corresponded to a moderate effect size (Cohen's d ≈ 0.6). No individual demographic or pre-existing academic variables were available from the anonymized institutional datasets, precluding formal assessment of baseline comparability between cohorts.

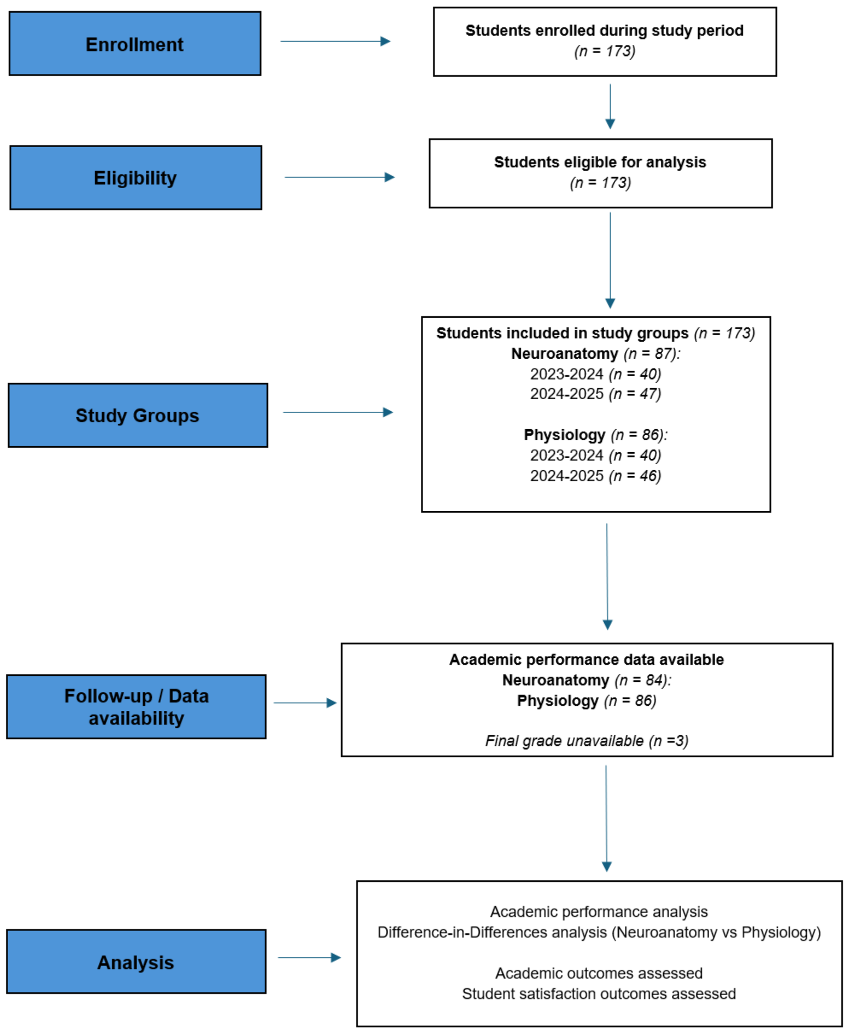

Figure 1. Flow diagram of cohort selection, data availability, and educational outcome assessment. *, Student satisfaction analyses were performed using available institutional survey data, and sample sizes varied across domains according to survey completion and availability of comparable items between academic years.

3.2. Global Academic Performance in Neuroanatomy: Flipped vs. Traditional

Academic performance was analyzed according to the type of assessment. High-stakes evaluation was represented by the final examination score, whereas low-stakes evaluation included continuous assessment components such as seminars, practical sessions, and individual coursework (table 2). The mean examination score was 3.43 ± 1.54 in the 2023–2024 cohort and 2.53 ± 1.53 in the 2024–2025 cohort. For low-stakes assessment, the global continuous evaluation score was 1.53 ± 0.45 in 2023–2024 and 0.97 ± 0.30 in 2024–2025. When disaggregated, the mean score for seminars and practical sessions was 1.09 ± 0.13 in 2023–2024 and 0.97 ± 0.30 in 2024–2025. The component corresponding to individual coursework and tests was only available in the 2023–2024 cohort (0.77 ± 0.14, n=23). This assessment component was not included as a separate evaluative element in the 2024–2025 academic year and therefore could not be analyzed independently for that cohort. Due to the non-normal distribution of the variables, comparisons were performed using the Mann–Whitney U test. Statistically significant differences were observed between cohorts for high-stakes assessment ($U = 610.5$, $Z = -2.68$, $p = 0.007$), low-stakes assessment overall ($U = 355.0$, $Z = -4.98$, $p < 0.001$), and the seminars and practical component ($U = 700.0$, $Z = -2.05$, $p = 0.041$). In addition, the overall final grade was significantly higher in the 2023–2024 cohort compared to the 2024–2025 cohort ($U = 568.5$, $Z = -2.77$, $p = 0.006$).

Table 2. Neuroanatomy course: academic performance according to assessment type and cohort.

Variable	2023-2024 (n)	2024-2025 (n)	p
High-stakes			
Final Examination	3.43 ± 1.54 (n=40)	2.53 ± 1.53 (n=46)	0.007
Low-stakes (overall)			
Continuous assessment	1.53 ± 0.45 (n=40)	0.97 ± 0.30 (n=47)	<0.001
Seminars and practicals	1.09 ± 0.13 (n=40)	0.97 ± 0.30 (n=47)	0.041
Coursework and tests	0.77 ± 0.14 (n=23)	Not available	
Overall outcome			
Final grade	4.44 ± 2.33 (n=39)	3.02 ± 2.21 (n=45)	0.006

*, Data are presented as mean ± standard deviation. P-values were obtained using the Mann–Whitney U test. High-stakes assessment corresponds to final examination scores. Low-stakes assessment includes continuous evaluation components such as seminars, practical sessions, and coursework.

3.3. Student Satisfaction and Perceived Teaching Quality

Because the institutional surveys used different scoring systems across academic years (1–7 in 2023–2024 and 1–5 in 2024–2025), raw satisfaction scores were not considered directly comparable. Therefore, all analyses were based on normalized scores transformed to a common 0–100 scale. After normalization, higher satisfaction levels were observed in 2024–2025, with mean values increasing from 77.98 ± 5.35 in 2023–2024 to 88.61 ± 3.65 in 2024–2025, and median values of 78.08 and 87.25, respectively. Because the institutional surveys differed between academic years in terms of scale structure and item composition, domain harmonization required the grouping of conceptually similar items and score normalization. Consequently, comparisons of student satisfaction across cohorts should be interpreted cautiously, as some observed differences may reflect characteristics of the survey instruments themselves rather than true differences in educational outcomes. Overall, after score normalization, students reported a more favorable perception of teaching quality in the 2024–2025 cohort. However, given the differences in survey design, scale structure, and item composition between academic years, these findings should be considered exploratory and cannot be attributed solely to differences in teaching methodology.

3.4. Comparison of Common Survey Items Between Academic Years

When survey items were grouped into conceptual domains, a consistent decline in student satisfaction was observed across most domains in the 2024–2025 academic year (table 3). After normalization of Likert-scale scores to a 0–100 range, statistically significant differences were found in Clarity and Structure ($p < 0.001$), with higher scores in the 2024–2025 academic year, indicating improved perceived clarity of teaching and course structure. Domains related to Methodology and Teaching Quality ($p = 0.064$) and Engagement and Participation ($p = 0.060$) showed a consistent trend toward higher scores in the 2024–2025 cohort, although these differences did not reach statistical significance. No significant differences were observed in Support and Accessibility ($p = 0.121$) or Relationship and Professionalism ($p = 0.121$). The domain of Motivation and Inspiration could not be statistically compared due to the absence of data in one of the groups. Overall, these findings suggest that student satisfaction tended to be higher in the 2024–2025 academic year across several domains, particularly those related to clarity, methodology, and engagement.

Table 3. Student satisfaction by domain across academic years.

Domain	2023-2024	2024-2025	p
Clarity and Structure	75.92 ± 3.38	89.38 ± 4.35	<0.001
Methodology and Teaching Quality	75.11 ± 5.85	89.38 ± 4.07	0.064
Engagement and Participation	69.50 ± 2.30	82.25 ± 0.71	0.060
Support and Accessibility	80.67 ± 5.10	92.75 ± 4.30	0.121
Motivation and Inspiration	71.44 ± 2.12	Not available	-
Relationship and Professionalism	72.0 ± 1.90	93.25 ± 1.75	0.121
Organization	89.0 ± 4.80	91.12 ± 5.13	0.502

*, Data are presented as normalized domain scores (0–100 scale) derived from institutional teaching evaluation surveys. Means and standard deviations reflect the distribution of item-level scores within each domain. Comparisons between academic years were performed using the Mann–Whitney U test. The domain Motivation and Inspiration consisted of three survey items available only in the 2023–2024 questionnaire. Because no conceptually equivalent items were present in the 2024–2025 survey, this domain could not be statistically compared and is presented for descriptive purposes only.

3.4.1. Teaching Clarity and Structure

After normalization, the domain of Clarity and Structure showed significantly higher scores in the 2024–2025 academic year compared to 2023–2024 ($p < 0.001$). This domain encompassed key elements such as clarity of explanations, definition of learning objectives, and overall organization of teaching sessions. Although higher scores were observed in the 2024–2025 cohort, these findings should be interpreted cautiously because the institutional surveys differed between academic years in terms of scale structure, item wording, and domain composition, requiring subsequent harmonization and score normalization. Therefore, the observed differences may reflect both variations in student perceptions and characteristics of the survey instruments themselves. Interestingly, this pattern contrasted with the decline in academic performance observed between cohorts, highlighting a possible dissociation between perceived teaching quality and objective educational outcomes.

3.4.2. Methodology and Student Engagement

The domains of Methodology and Teaching Quality and Engagement and Participation showed a consistent trend toward higher scores in the 2024–2025 cohort, although these differences did not reach statistical significance (p = 0.064 and p = 0.060, respectively). These domains reflect students' perceptions of teaching strategies, learning activities, and their level of active involvement during sessions. Therefore, these findings should be considered exploratory and interpreted cautiously. While the direction of the observed differences may indicate a tendency toward more favorable evaluations in the 2024–2025 academic year, the lack of statistical significance and the differences between survey instruments preclude definitive conclusions regarding the effect of teaching methodology.

3.4.3. Perceived Support and Teacher–Student Interaction

No significant differences were observed in the domains of Support and Accessibility and Relationship and Professionalism between academic years (p = 0.121 for both). These domains assess the availability of the instructor and the quality of teacher–student interaction. The stability of these dimensions suggests that interpersonal aspects of teaching and perceived support remained consistent across cohorts.

3.4.4. Motivation and Professional Impact

The domain of Motivation and Inspiration could not be statistically compared between academic years because the three items comprising this domain ("The teacher motivates and inspires me", "The teacher stimulates curiosity", and "The teacher encourages professional development") were only available in the 2023–2024 survey, and no conceptually equivalent items were included in the 2024–2025 survey. Consequently, this domain was excluded from formal comparative analyses and is reported only descriptively for the 2023–2024 cohort. This limitation reflects differences in survey structure rather than a true absence of motivational factors and therefore precludes any conclusions regarding temporal changes in this domain.

3.4.5. Organization and Course Logistics

The domain of Organization did not show significant differences between academic years after normalization (p = 0.502). This domain included items related to punctuality, use of teaching resources, and logistical aspects of the course. These findings suggest that perceived course organization remained stable across cohorts.

3.4.6. Summary of Satisfaction Trends

Overall, normalized student satisfaction scores were higher in the 2024–2025 academic year across several domains. The largest differences were observed in the domain of Clarity and Structure, while higher scores were also observed in the domains of Methodology and Teaching Quality and Engagement and Participation, although these differences did not reach statistical significance. In contrast, the domains of Support and Accessibility, Relationship and Professionalism, and Organization showed smaller differences between cohorts. The Motivation and Inspiration domain could not be compared because corresponding items were unavailable in the 2024–2025 survey.

3.5. Control Group (Physiology) and Difference-in-Differences Analysis

To control for potential cohort-related differences, academic performance in a parallel course (physiology), delivered using a stable teaching methodology across both academic years, was analyzed (table 4). A total of 86 students were included in the physiology cohort, with 40 students in the 2023–2024 academic year and 46 in the 2024–2025 academic year. Final grade data were available for 40 and 32 students, respectively. The mean final grade in physiology was 6.20 ± 1.27 in the 2023–2024 cohort and 3.39 ± 2.98 in the 2024–2025 cohort. A statistically significant difference was observed between cohorts (U = 313.5, Z = -3.70, $p < 0.001$). When analyzing the assessment structure, the mean examination score was 3.89 ± 1.19 in 2023–2024 and 3.06 ± 1.39 in 2024–2025, showing a significant difference (U = 601.0, Z = -2.76, p = 0.006). For low-stakes assessment, the overall continuous evaluation score was 2.31 ± 0.40 in 2023–2024 and 1.77 ± 0.71 in 2024–2025, with a statistically significant difference between cohorts (U = 452.0, Z = -3.72, $p < 0.001$). When disaggregated, no statistically significant differences were observed in the seminars and practical component (U = 702.0, Z = -1.09, p = 0.275) or in coursework and tests (U = 391.0, Z = -1.09, p = 0.275).

To assess whether the observed changes in academic performance in neuroanatomy were attributable to the teaching methodology rather than cohort effects, a difference-in-differences (DiD) analysis was conducted using physiology as a control group.

A linear regression model was fitted including academic year, subject (neuroanatomy vs. physiology), and their interaction term. The interaction term, representing the difference-in-differences estimator, showed a negative coefficient (β = -1.387; 95% CI [-2.814, 0.039]; p = 0.057), suggesting a greater decline in neuroanatomy performance compared to physiology between academic years. Although the interaction term did not reach conventional statistical significance, the magnitude and direction of the estimate may indicate a potential association that warrants cautious interpretation. Given the limited sample size and the possibility of type II error, these findings should not be interpreted as definitive evidence of the absence of an effect. Nevertheless, causal inferences remain inappropriate because of the observational nature of the study and the inability to formally verify the parallel trends assumption.

Academic year was independently associated with lower performance (β = -1.419, p = 0.004), supporting the presence of a cohort effect affecting both subjects. In addition, physiology showed higher overall grades compared to neuroanatomy (β = 1.759, 95% CI [-2.386, -0.452]; p = 0.001). Overall, although the direction of the effect suggests a possible benefit of the flipped classroom, the observed differences cannot be conclusively attributed to the change in teaching methodology after adjusting for cohort-related factors.

Table 4. Physiology course (control group): academic performance according to assessment type and cohort.

Variable	2023-2024 (n)	2024-2025 (n)	p
High-stakes			
Final Examination	3.89 ± 1.19 (n=40)	3.06 ± 1.39 (n=46)	0.006
Low-stakes (overall)			
Continuous assessment (overall)	2.31 ± 0.40 (n=40)	1.77 ± 0.71 (n=43)	<0.001
Seminars and practicals	1.27 ± 0.10 (n=38)	1.13 ± 0.27 (n=43)	0.275
Coursework and tests	1.13 ± 0.15 (n=39)	1.15 ± 0.20 (n=24)	0.275
Overall outcome			
Final grade	6.20 ± 1.27 (n=40)	3.39 ± 2.98 (n=32)	<0.001

*, Data are presented as mean ± standard deviation. P-values were obtained using the Mann–Whitney U test. High-stakes assessment corresponds to final examination scores. Low-stakes assessment includes continuous evaluation components such as seminars, practical sessions, and coursework.

A difference-in-differences analysis of student satisfaction was conducted using physiology as a control group. The interaction term suggested differences in satisfaction across methodologies, particularly in engagement, although results should be interpreted cautiously given sample size and survey heterogeneity. This effect reached statistical significance in the domain of Engagement and Participation (β = -7.17; 95% CI [-10.893, -3.440]; p = 0.014), suggesting that students exposed to the flipped model reported significantly lower engagement compared to the control condition. Given that the flipped classroom was implemented in 2023–2024, this finding is consistent with the higher satisfaction observed in the 2024–2025 cohort. A similar negative trend was observed in Clarity and Structure (β = -6.38; 95% CI [-13.780, 1.022]; p = 0.086), although this did not reach statistical significance (table 5). No significant effects were observed in other domains, likely due to limited sample size.

Table 5. Difference-in-Differences Regression Models: Effect Estimates, 95% Confidence Intervals, and Statistical Significance for Academic Performance and Student Satisfaction Outcomes.

Outcome	β coefficients	95% CI	p
Final grade	-1.387	-2.814 to 0.039	0.057
Engagement	- 7.17	-10.893 to -3.440	0.014
Clarity	-6.38	-13.780 to 1.022	0.086

*, β coefficients represent the interaction term (subject × academic year) obtained from the Difference-in-Differences regression models. Negative β values indicate a greater decline (or a smaller increase) in neuroanatomy relative to physiology between academic years. Results are presented as β coefficients with 95% confidence intervals (95% CI) and corresponding p-values. Statistically significant results are highlighted at $p < 0.05$. DiD, Difference-in-Differences; CI, confidence interval.

4. Discussion

This quasi-experimental study compared the academic performance and student satisfaction associated with two teaching methodologies in neuroanatomy across two consecutive academic years, using a parallel physiology course as a control group and a difference-in-differences (DiD) approach to account for inter-cohort variability.

A significant decline in academic performance was observed in neuroanatomy between the 2023–2024 (flipped classroom) and 2024–2025 (traditional lecture-based teaching) academic years across all assessment components, including final examination scores, continuous assessment, and overall final grades. A similar decline occurred in the physiology control group, where the teaching methodology remained unchanged. This parallel trend suggests the presence of cohort-related factors influencing academic performance. Although the DiD analysis showed a slightly greater decline in neuroanatomy (β = -1.387; p = 0.057), the interaction term did not reach conventional statistical significance. Therefore, while the findings do not provide sufficient evidence to demonstrate an independent effect of teaching methodology on academic performance, the possibility of a modest effect cannot be excluded and should be explored in larger studies.

In contrast, student satisfaction showed a different pattern. After normalization of scores to a 0–100 scale, overall satisfaction was higher in the 2024–2025 academic year (traditional lecture-based teaching). The DiD analysis confirmed that these differences were not explained solely by cohort effects. Specifically, the interaction term for the Engagement and Participation domain was statistically significant (β = -7.17; 95% CI [-10.893, -3.440]; p = 0.014), indicating that the flipped classroom model was associated with significantly lower student-reported engagement compared to the traditional lecture-based teaching approach, after adjusting for secular trends. A similar (non-significant) trend was observed in the Clarity and Structure domain (β = -6.38; p = 0.086).

Several factors may help explain the lower engagement scores observed in the flipped classroom cohort. Active-learning methodologies often require greater student preparation and self-directed learning before classroom sessions, which may increase perceived workload and cognitive demands. In addition, students accustomed to traditional lecture-based teaching formats may initially experience resistance when confronted with more active and participatory learning strategies. Differences in expectations regarding the roles of students and instructors may also influence satisfaction ratings. Consequently, lower engagement scores should not necessarily be interpreted as evidence of reduced educational value, but rather as a reflection of the challenges associated with adaptation to active-learning environments.

These findings highlight a clear dissociation between student satisfaction and academic performance. While the traditional lecture-based teaching methodology implemented in 2024–2025 was perceived more favorably (particularly in engagement and clarity), this did not translate into better academic outcomes. Students may prefer methodologies that feel clearer or less demanding, even if these do not necessarily lead to superior knowledge acquisition or examination performance.

The results align with the heterogeneous literature on flipped classroom methodologies in medical education. While some studies report benefits in satisfaction or performance, many lack adequate controls for cohort variability. The present study adds to the field by incorporating a concurrent control group and a DiD framework, providing a more robust assessment of associations between teaching approaches and educational outcomes while partially controlling for cohort-related variability.

The main strengths of this study include its quasi-experimental design incorporating a parallel control subject and the use of a Difference-in-Differences (DiD) analytical framework, which provides a more rigorous approach than simple before–after comparisons. Nevertheless, several limitations should be acknowledged. First, the study was conducted at a single institution and included a relatively small sample size, particularly for some student satisfaction domains. Second, the normalization and harmonization of survey instruments originally based on different scoring systems may have introduced some measurement variability. Furthermore, as a non-randomized study, residual confounding cannot be completely excluded despite the use of a concurrent control group.

An additional limitation concerns the parallel trends assumption underlying DiD analyses. Because only two academic years were available and no pre-intervention observations existed beyond the study period, it was not possible to formally assess whether neuroanatomy and physiology followed similar trends before the implementation of the teaching change. Therefore, the DiD findings should be interpreted as providing a framework to partially account for cohort-related variability rather than as definitive evidence of a causal effect of teaching methodology. Future multicenter studies with longer observation periods and multiple pre-intervention measurements would help strengthen causal inference and improve the generalizability of these findings.

Furthermore, the comparability of student satisfaction measures across academic years represents an important source of uncertainty. Although survey scores were normalized and conceptually similar items were grouped into common domains, the original questionnaires differed in scale format, item wording, and domain composition. Some domains contained incomplete information, and the "Motivation and Inspiration" domain could not be evaluated in both cohorts. Consequently, changes observed in satisfaction outcomes may partly reflect differences in the measurement instruments rather than true differences associated with teaching methodology. Therefore, these findings should be interpreted cautiously and considered exploratory.

Finally, no formal sample size calculation was performed because all available students from the study period were included. Consequently, the relatively small sample size, particularly for some satisfaction domains, may have limited statistical power and increased the risk of type II error. This limitation is especially relevant for several findings that approached statistical significance, including the main Difference-in-Differences analysis ($p = 0.057$) and some satisfaction domains. Therefore, the absence of statistically significant differences should not necessarily be interpreted as evidence of no effect. Future studies should incorporate systematic reporting of effect size measures across all outcomes to facilitate interpretation of the practical significance of educational interventions.

In conclusion, in this real-world educational setting, neither methodology demonstrated superiority in academic performance after adjusting for cohort effects. However, students exposed to the teaching approach implemented in 2024–2025 reported higher satisfaction and engagement, although causal inferences regarding the effect of methodology should be interpreted with caution. These results underscore the importance of using controlled designs in educational research and evaluating interventions based on both objective academic outcomes and subjective student experience.

5. Conclusions

- In this quasi-experimental study, no teaching methodology demonstrated superiority in academic performance after adjustment for cohort effects using a difference-in-differences approach. Although academic performance declined between academic years, this pattern was observed in both neuroanatomy and physiology, suggesting the influence of cohort-related factors. Therefore, the available data do not allow the observed changes in academic performance to be conclusively attributed to the teaching methodology alone. After score normalization, higher levels of student satisfaction and engagement were observed in the 2024–2025 cohort; however, these findings should be interpreted cautiously given differences in survey design and the limitations inherent to the quasi-experimental design.
- The findings reveal a dissociation between perceived teaching quality and objective academic outcomes, indicating that greater student satisfaction does not necessarily translate into improved examination performance. These results emphasize the importance of evaluating educational interventions using controlled designs that incorporate both objective and subjective educational measures.
- Educational innovations in medical education should not be assessed solely through student satisfaction metrics. Controlled methodological approaches are essential to determine the true academic impact of teaching strategies, particularly in complex disciplines such as neuroanatomy. Future research should incorporate larger sample sizes, longitudinal follow-up, and objective measures of learning outcomes to strengthen causal inference and improve understanding of the educational effectiveness of different teaching methodologies.

Funding: There was no funding.

Declaration of conflict of interest: The authors declare that they have no conflict of interest.

Author contributions: J.F.G.G. conceived and designed the study, collected the data, performed the statistical analysis, interpreted the results, and drafted the manuscript. A.N.G.H. contributed to study supervision, interpretation of the findings, and critical revision of the manuscript for important intellectual content. All authors reviewed and approved the final version of the manuscript.

6. References.

1. Chen F, Lui AM, Martinelli SM. A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education*. **2017,** 51(6), 585-597. https://doi.org/10.1111/medu.13272
2. Hew KF, Lo CK. Flipped classroom improves student learning in health professions education: a meta-analysis. *BMC Medical Education*. **2018,** 18(1), 38. https://doi.org/10.1186/s12909-018-1144-z
3. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. Active learning increases student performance in science, engineering, and mathematics. *Psychological and Cognitive Sciences*. **2014,** 111(23), 8410-5. https://doi.org/10.1073/pnas.1319030111
4. Prober CG, Heath C. Lecture halls without lectures — a proposal for medical education. *New England Journal of Medicine*. **2012,** 366(18), 1657–1659. https://doi.org/10.1056/nejmp1202451
5. Chen KS, Monrouxe L, Lu YH, Jenq CC, Chang YJ, Chang YC, Chai PY. Academic outcomes of flipped classroom learning: a meta-analysis. *Medical Education*. **2018,** 52(9), 910–24. https://doi.org/10.1111/medu.13616
6. O'Flaherty J, Phillips C. The use of flipped classrooms in higher education: a scoping review. *The Internet and Higher Education*. **2015,** 25, 85–95. https://doi.org/10.1016/j.iheduc.2015.02.002

7. Uttl B, White CA, Wong Gonzalez D. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*. **2017,** 53, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007
8. Braga M, Paccagnella M, Pellizzari M. Evaluating students' evaluations of professors. *Economics of Education Review*. **2014,** 41, 71–88. http://dx.doi.org/10.1016/j.econedurev.2014.04.002
9. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. *Boston: Houghton Mifflin*; **2002**. https://iaes.cgiar.org/sites/default/files/pdf/147.pdf
10. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. **2014,** 312(22), 2401-2. https://doi.org/10.1001/jama.2014.16153