A Data-Driven Institutional Study of Academic Performance: Exploring Course, Structure, and Environment in Higher Education

by

Arif Shakil
18266002

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
M.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Arif Shakil
18266002

Approval

The thesis/project titled "A Data-Driven Institutional Study of Academic Performance: Exploring Course, Structure, and Environment in Higher Education" submitted by

1. Arif Shakil(18266002)

Of Summer 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science on 19th October, 2025.

Examining Committee:

Supervisor:
(Member)

Sadia Hamid Kazi

Associate Professor
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
BRAC University

Internal:
(Member)

Muhammad Iqbal Hossain

Associate Professor
Department of Computer Science and Engineering
BRAC University

External:
(Member)

Mohammad Zahidur Rahman

Professor
Department of Computer Science and Engineering
Jahangirnagar University

Thesis Coordinator:
(Member)

Md. Sadek Ferdous

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi

Associate Professor
Department of Computer Science and Engineering
BRAC University

Ethics Statement

Strict adherence to research ethics was maintained throughout the study. No personally identifiable information was accessed or stored. Institutional approval was obtained from BRAC University. All data was used exclusively for academic research and stored securely to prevent misuse.

Glossary

RS (Residential Semester): A compulsory, immersive academic term held at BRAC University's residential campus. It emphasizes collaborative learning, community living, and focused coursework, distinct from the main campus setting.

Main Campus: The primary academic campus where most courses are conducted in a regular, non-residential setting.

Delivery Modes: The method of course instruction. Classified as *Offline (in-person)*, *Online (virtual)*, and *Hybrid (a combination of online and in-person elements)*.

Year–Semester Code (yyyyx): A chronological identifier encoding semester sequence, where Spring=1, Summer=2, Fall=3 (e.g., 20213 = Fall 2021).

COVID Phases: Categorized for temporal comparison:

- **Pre-COVID:** Up to Fall 2019 (20193)
- **During-COVID:** Spring 2020–Summer 2021 (20201–20212)
- **Post-COVID:** Fall 2021 (20213) onward

[6pt]
CGPA (Cumulative Grade Point Average): The overall weighted average GPA across all completed courses, computed on a 4.0 scale.

GPA (Grade Point Average): The mean performance for a single semester or course, measured on a 4.0 scale.

Class Size: The total number of students enrolled in a specific course section during a particular semester.

Impact Score: A metric combining the correlation between course GPA and CGPA with course enrollment frequency, used to identify high-leverage courses.

Abstract

Student performance reflects the intersection of academic, structural, and environmental factors rather than individual effort alone. This study conducts a large-scale, data-driven analysis of institutional records from BRAC University to examine how these factors shape outcomes across time, delivery modes, and disciplines. Nine hypotheses were tested, encompassing the effects of online, hybrid, and inperson learning environments; the COVID-19 pandemic; course-level contributions to CGPA; prerequisite–core alignment; class size; Residential Semester (RS) contexts; high-failure course patterns; and inter-departmental performance differences. Using Pearson correlation, regression, ANOVA, Kruskal–Wallis, Welch's t-tests, and mixed-effects models, the study identifies clear structural trends: RS participation strongly correlates with higher and more consistent GPA; class size shows weak, context-dependent effects; and persistent high-failure rates cluster in STEM gateway courses. Departments differ systematically—humanities and law programs maintain higher GPAs, while technical disciplines show greater variance due to assessment rigor and prerequisite dependency. These findings reveal that academic outcomes are institutionally patterned, not random. They underscore the need for data-informed curriculum design, departmental benchmarking, and early-risk intervention frameworks to promote equity, quality, and resilience in higher education.

Keywords: Academic Performance, Class Size, Prerequisite Courses, Course Failure Rates, Educational Data Mining

Dedication

To the students of BRAC University, whose struggles and successes gave meaning to this research. To the teachers and mentors who continue to shape lives through patience and dedication. And to the academic environment that inspired this exploration into the many forces behind student performance.

Acknowledgment

Firstly, all praise is due to the Almighty Allah, whose mercy and guidance enabled me to complete this work.
I would like to express my deepest gratitude to my Supervisor, Dr. Sadia Hamid Kazi miss and Co-Supervisor, Md. Golam Rabiul Alam Sir, for their invaluable support, advice, and encouragement throughout this research. Their insights and direction were instrumental in shaping the study and keeping it on track.
I would like to extend special thanks to Joydhriti Choudhury and Faisal Ahmed for their assistance in collecting data. I am also deeply grateful to Adria Binte Habib and Md. Sabbir Ahmed for their guidance and continuous encouragement during the preparation of this thesis.
I am equally thankful to my peers, whose discussions, constructive feedback, and help with validations contributed to the clarity and depth of this work. Finally, I extend my heartfelt appreciation to my family for their patience, constant support, and prayers, which sustained me through the challenges of this journey. Without their encouragement, this thesis would not have been possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Student success in a university setting is shaped by multiple forces that interact over time. These include course and program structure, instructional quality, class size and format, and the broader learning environment. The COVID-19 pandemic accelerated a shift to online delivery, revealed gaps in digital access and readiness, and introduced stressors that changed how students learn [54], [61]. At the same time, long-standing structures such as prerequisite sequences, course clusters, and class size continue to influence performance in ways that vary across disciplines.
BRAC University provides a valuable context for examining these dynamics. Its layered academic environment spans foundational courses, advanced tracks, and multiple delivery modes, including the **Residential Semester (RS)** and main campus settings. The RS is physically and organizationally distinct, which raises an important question: does student performance change when location, schedule, and delivery constraints differ from the main campus experience? Beyond location and mode, structural features such as class size distributions, prerequisite–core alignment, and the presence of at-risk courses are central. Non-academic elements such as access to different environment and well-being resources may also support achievement by promoting balance and resilience. This thesis adopts a longitudinal, data-driven approach to understand how academic, structural, and environmental factors interact. The aim is to provide evidence that can inform curriculum planning, instructional design, advising, and policy.

1.1 Research Motivation

Understanding why students succeed or struggle is both an academic and an equity concern. Metrics such as grades, CGPA, withdrawals, and repeat attempts signal how well systems support diverse learners. The RS context at BRAC University, variations in class size, differences in delivery mode, and prerequisite scaffolding offer a natural set of contrasts. Prior work suggests that class size can shape motivation and performance, often through reduced individual attention and increased anonymity in larger classes [51], [53]. Delivery method also matters, with motivation and engagement differing by online versus face-to-face contexts [67]. Evidence on prerequisites indicates that mastery can support later success, though effects may vary by discipline and may be partially mediated by broader indicators such as overall GPA or academic standing [41], [46], [85]. The pandemic period provides a temporal lens that helps distinguish temporary disruptions from structural patterns that persist [54], [61].

1.2 Research Objectives

Understanding the determinants of student performance requires a holistic examination of the institutional, structural, and contextual variables that shape academic outcomes. Rather than focusing on isolated factors, this study adopts a systems-level approach, integrating course-level, program-level, and university-level perspectives. The objectives below are designed to systematically uncover how delivery modes, class sizes, campus environments, and curricular linkages collectively influence performance metrics such as GPA and CGPA. These insights aim to inform data-driven policy, curriculum design, and student support frameworks at the institutional level. This study investigates structures and contexts that influence student performance. The specific objectives are:

1. Compare outcomes across **online, in-person, and hybrid** delivery modes, including the pandemic period, to assess how instructional format influences GPA trends.

2. Identify **individual courses** that disproportionately influence CGPA, determining whether performance in those courses tends to raise or lower overall academic standing.

3. Test whether strong **prerequisite performance** predicts success in corresponding core or advanced courses.

4. Examine the academic impact of the **COVID-19 pandemic period** on student performance across semesters and programs.

5. Evaluate the relationship between **class size** and academic outcomes across disciplines.

6. Compare performance across **main campus** and **Residential Semester (RS)** contexts, emphasizing differences in **location, learning environment, and living conditions**.

7. Flag **courses with consistently high failure or withdrawal rates** and explore the **structural factors**—such as prerequisite alignment, campus environment, faculty dynamics, and prior student performance—underlying these patterns.

8. Analyze **program-level and department-level** performance differences linked to curriculum structure, assessment design, and faculty distribution.

Each of these objectives is rooted in a commitment to educational equity, academic quality, and institutional self-awareness. Together, the objectives aim to provide a comprehensive and context-aware understanding of what drives student performance in a modern, dynamic university environment like BRAC University.

1.3 Research Questions and Hypotheses

To explore the structural and contextual factors shaping student academic performance, this study is guided by nine focused research questions and hypotheses. Each hypothesis corresponds to an empirically tested component in the results chapter, reflecting BRAC University's institutional dynamics across delivery modes, course structures, and learning environments.
RQ1. How do **online, in-person, and hybrid delivery modes** affect student academic performance across programs and semesters? **H1.** Delivery mode has a statistically significant effect on student performance, with in-person formats generally associated with higher GPA, though variation depends on course type, instructional design, and digital readiness.

RQ2. Which **individual courses** have the strongest correlation with cumulative CGPA, and how much do they contribute to overall academic outcomes? **H2.** Certain high-credit or conceptually demanding courses, particularly in programming and quantitative areas, exert disproportionate influence on CGPA and act as key performance drivers within programs.

RQ3. Does strong **prerequisite performance** predict higher achievement in subsequent core or advanced courses? **H3.** Students who achieve strong grades in prerequisite courses, especially in foundational programming and mathematics, are more likely to perform well in their corresponding core courses, indicating effective knowledge transfer and scaffolding.

RQ4. Did the **impact of the COVID-19 pandemic** on academic performance vary **across departments or disciplines**?
H4. The pandemic's academic impact differed significantly among departments, with some disciplines demonstrating stronger adaptability and resilience, while others—particularly those relying on laboratory or design-based instruction—experienced greater performance declines.

RQ5. How did **student performance change across different phases** of the pandemic (**Pre-COVID, During-COVID, Post-COVID**)?
H5. Student GPA varied significantly across pandemic phases, showing a performance spike during the remote-learning period followed by a partial normalization post-pandemic, indicating temporal adaptation patterns in academic outcomes.

RQ6. Does **class size** correlate with student performance, and how do these effects vary across disciplines? **H6.** Smaller class sizes are positively correlated with higher average GPA, suggesting that individualized attention and manageable student–faculty ratios contribute to better learning outcomes.

RQ7. How does academic performance differ between the **Residential Semester (RS)** and main campus environments? **H7.** RS semesters are associated with statistically significant performance differences compared to main campus semesters, driven by variations in location, scheduling, residential engagement, and focused resource access.

RQ8. Which courses exhibit **consistently high failure or withdrawal rates**, and what structural patterns explain these trends? **H8.** Persistently difficult courses share identifiable features such as high conceptual load, prerequisite misalignment, intensive assessments, or high instructor turnover, making them priority targets for academic support and redesign.

RQ9. Are there **programs or departments** where students consistently perform better or worse, and what factors contribute to these patterns? **H9 (Integrative Hypothesis).** Statistically significant performance variation exists across programs and departments, reflecting the combined influence of delivery mode, curriculum structure, assessment design, and environmental factors identified across H1–H8.

1.4 Contributions

This section summarizes how the thesis advances both understanding and practice. The work brings together factors that are usually studied separately, such as delivery mode, period effects, class size, prerequisite scaffolding, course clusters, and the Residential Semester environment, into one program-aware, time-aware, and course-aware framework. Using multi-year, section-level data from BRAC University in a resource-constrained context, the study pairs careful statistical analysis with actionable diagnostics that departments and program offices can use for curriculum design, sectioning, advising, and targeted support. The main contributions are outlined below.

- **Pandemic-period analysis** that situates delivery-mode effects in time and across disciplines.
- **Learning environment insights** comparing online, hybrid, and in-person performance to guide delivery models.
- **Course-level influence estimates** that quantify how individual and prerequisite courses shape CGPA.
- **Cluster-level perspectives** that emphasize cumulative pathways rather than isolated courses.
- **Class size evidence** across departments to inform sectioning and staffing.
- **Non-academic factors** that broaden the discussion to student well-being and balance.
- **Actionable diagnostics** that identify at-risk courses for redesign, supplemental instruction, or advising.
- **Integrative insight (H9)** revealing cross-program and inter-departmental performance patterns that emerge from the combined influences of delivery mode, curriculum design, and environmental factors across H1–H8.

1.5 Expected Outcomes

This research is designed not merely to explore academic questions, but to generate concrete, institutionally relevant outcomes that can inform both immediate interventions and long-term academic strategy. Through careful examination of course, semester, and program data at BRAC University, the study seeks to produce a layered understanding of student performance as shaped by structural and contextual variables. The expected outcomes of this research include the following:

- Comparative profiles of **RS vs main campus** performance.
- Effect estimates for **class size** by department and course type.
- A ranked list of **high-impact courses** and their estimated contribution to CGPA.
- Evidence on the **predictive value of prerequisites**.
- Mode-wise analyses across **online, in-person, and hybrid** delivery.
- Program-level **performance maps** highlighting structural differences.
- A diagnostic set of **at-risk courses** with likely causes.
- A **scalable monitoring framework** for ongoing policy feedback.

By transforming performance data into actionable insights rather than abstract conclusions, this research aims to equip BRAC University with the tools to refine its academic structures, strengthen educational equity, and respond effectively to both long-term trends and sudden disruptions. In doing so, it contributes to a deeper, evidence-based understanding of what drives student success. Insights that can guide not only institutional policy but also the everyday practices that shape learning outcomes.

1.6 Thesis Structure

This thesis is organized into five chapters, each building upon the last to form a coherent investigation into the structural factors influencing student performance at BRAC University. The structure is designed to gradually transition from problem identification and contextual framing to empirical analysis and actionable insights.

Chapter 2: Literature Review The second chapter surveys prior research on academic performance, curriculum design, class size, online versus offline learning, prerequisite effectiveness, and institutional learning environments. It draws from both international and local sources, identifying gaps in current knowledge and situating this study within the broader academic discourse.

Chapter 3: Methodology This chapter details the dataset used, drawn from multiple semesters and departments of BRAC University since the start of the journey of the institution, and the techniques applied for data preprocessing, transformation, and filtering. It outlines the statistical tools and models used, including correlation analysis, regression, group comparisons, and trend analysis across semesters and course structures.

Chapter 4: Results and Discussion Here, the study presents findings corresponding to each research question and hypothesis. It includes visualizations, tables, and comparative analyses, along with interpretive commentary linking statistical patterns to institutional practices. This chapter also reflects critically on unexpected results, discipline-specific deviations, and systemic implications.

Chapter 5: Conclusion and Future Work The final chapter summarizes the key insights and contributions of the study. It reflects on the institutional relevance of the findings and outlines practical recommendations for curriculum design, academic advising, and semester planning. The chapter also suggests directions for future research, including extensions of the current framework and possible qualitative follow-ups.

Chapter 2

Related Work

Understanding student performance in a university setting goes far beyond the simple question of who studies more or who works harder. Academic outcomes are shaped by an interplay of structural, environmental, and curricular factors. These include the mode of course delivery. For example, whether learning takes place entirely in a physical classroom, entirely online, or in some hybrid form. Each mode carries distinct benefits and limitations depending on context. Timing also plays a crucial role. For example, semesters that fell during the COVID-19 pandemic were influenced not only by changes in teaching format but also by altered assessment policies, varying levels of student engagement, and the wider emotional and economic stresses of that period.
On the other hand, there may be the influence of class size, which can affect the level of individual attention students receive, the amount and quality of feedback provided, and the degree of peer interaction possible in a given course. The scaffolding effect of prerequisite courses is another key component, as early foundational learning can either prepare students for success in advanced subjects or, if weak, create long-lasting gaps in understanding. Beyond the purely academic structure, the learning environment itself plays a significant role. BRAC University's Residential Semester (RS), for example, creates a concentrated, immersive experience with smaller cohorts, fewer distractions, and increased peer bonding, which may alter student performance patterns in ways that differ from the main campus.
Finally, not all courses contribute equally to a student's cumulative GPA (CGPA). Some courses that are large, required, or conceptually central have disproportionate weight in determining long-term academic trajectories. The combination of these factors forms a complex ecosystem in which institutional policies, teaching practices, and student behaviors interact. This chapter examines recent empirical evidence on each of these dimensions, with particular attention to findings from the Global South and resource-constrained contexts. It also frames how these insights directly inform the statistical modelling strategies employed in this thesis, ensuring that the analytical approach remains grounded in both international research and the specific realities of BRAC University.

2.1 Delivery Mode: Online vs. Offline vs. Blended

COVID-19 forced a rapid turn to online medium. Some studies reported small performance gains during lockdown, often attributed to flexibility and fewer commuterelated costs [44]. Yet post-pandemic evidence is more mixed: students value flexibility, but many still report weaker interaction and lower perceived learning support when everything moves online [69]. Large cross-country datasets point to demographic differences in engagement, with younger students or earlier academic levels sometimes responding more positively to online formats than seniors preparing for higher-stakes assessments [59].

Bangladesh fits the global pattern but with sharper constraints: unstable connectivity, device scarcity, and urban-rural gaps limited what online learning could practically deliver [55], [65], [78]. In BRAC University's case, these disparities were especially visible between main-campus students with ready access to buX and RS or rural students who often had to rely on mobile data or intermittent broadband, making synchronous participation uneven. Even when platforms were available (for example Zoom and LMS), student engagement depended on bandwidth, digital skills, and instructor adaptation [65], [71]. Meta-analyses of blended learning generally show small but consistent benefits over pure face-to-face or pure online, suggesting that mixing synchronous contact with asynchronous review materials can be a pragmatic middle path [64]. This study follows that nuance: instead of treating online versus offline as a single effect, the analysis compares outcomes program-wise and semester-wise to see where online helped, where it did not, and whether blended-like implementations during transition periods behaved differently.

Building on these pre-pandemic comparisons, recent empirical analyses underscore the persistent advantages of in-person and blended modes for achievement in STEM and biological sciences. For instance, [73] analyzed multi-institutional data and found in-person learning superior to fully online formats, with effect sizes up to 0.25 SD in GPA, recommending hybrid models for equity. Similarly, [76] reported that blended approaches yielded 10–15% higher marks in first-year biology courses than face-to-face alone, attributing gains to flexible digital scaffolding. In computer-science contexts, [79] observed comparable grades across modes but stronger peer support and belonging in blended settings, suggesting that structured attendance incentives can mitigate online isolation. These results reinforce this thesis's focus on mode-specific predictors and the nuanced role of delivery context in performance outcomes.

2.2 Course Performance and CGPA: Identifying Leverage Courses

While cumulative GPA aggregates performance across all coursework, certain courses, especially those that are mandatory, early in sequence, or high in enrollment, have a disproportionate impact on academic trajectories. Institutions often focus on improving overall CGPA without systematically identifying which specific courses exert the most influence [18]. Recent work in learning analytics advocates weighting course impact by both its correlation with CGPA and its enrollment reach, ensuring that interventions are targeted where they can benefit the largest number of students [31], [38].

Evidence from active-learning redesigns in high-enrollment STEM courses shows substantial gains in both average grades and pass rates, especially when changes are sustained over multiple semesters [47]. This approach has a double benefit: it can lift institutional performance metrics while simultaneously reducing equity gaps in high-risk courses. Conversely, neglecting leverage courses can allow persistent performance bottlenecks to remain unaddressed for years, leading to repeated cycles of student attrition or delayed graduation.

In this statistical study, the identification process begins with a simple yet informative metric: multiplying the Pearson correlation between course grade and CGPA by the logarithm of the number of students who have taken the course. For BRAC, this is especially pertinent because a handful of high-enrollment courses, such as ENG101, CSE110, and MAT110, are taken by thousands of students across programs, meaning small shifts in average performance can influence institutional CGPA at scale. This impact score provides an initial ranking of courses by practical leverage. High-scoring courses are then analyzed using multivariate regression models to control for potential confounders and to verify whether their observed influence holds when other factors are considered. The purpose is not to infer causation directly but to produce an evidence-based shortlist of courses where deeper pedagogical, curricular, or support interventions could yield the highest institutional returns.

While limited 2023–2025 work isolates leverage courses directly, emerging meta-trends from scaffolding studies suggest that foundational subjects disproportionately stabilize CGPA trajectories. For instance, [72] indirectly identified prerequisite mastery as a high-leverage factor for long-term academic stability. This evidentiary gap underscores the novelty of the present study's course-level regression approach to identifying leverage courses.

2.3 Prerequisites and Scaffolded Learning

The assumption that prerequisites lay a durable foundation for downstream academic success is widely accepted, yet empirical support is mixed once contextual variables are accounted for. In STEM education, multiple longitudinal studies have confirmed that early course performance is a strong predictor of later retention and achievement, particularly in cumulative-knowledge domains such as mathematics, physics, and programming [20], [39]. Large institutional datasets from engineering programs suggest that performance in advanced mathematics, such as calculus, and foundational physics courses correlates positively and often linearly with performance in subsequent engineering subjects [48].

However, the predictive strength of prerequisites is neither universal nor immune to erosion. In settings where instructional quality, course alignment, or student support structures are weak, even well-chosen prerequisites may fail to deliver their intended scaffolding effect. This is especially true in programming-intensive pathways in developing countries, where students may complete an introductory programming course without achieving sufficient mastery to handle more advanced problem-solving tasks [60].

The literature also hints at a possible over-reliance on prerequisites as gatekeeping mechanisms rather than developmental tools. At BRAC, foundational sequences such as CSE110 to CSE221 or MAT110 to MAT215 are assumed to prepare students for advanced work, but preliminary analysis suggests variation in how strongly these sequences predict later success across programs. If prerequisite courses are designed with overly compressed syllabi, misaligned assessments, or insufficient for-

mative feedback, they risk becoming hurdles rather than scaffolds. In this thesis, course-to-course relationships, for example CSE110 to CSE111, are tested using both correlation and regression models on large samples. This is followed by a program-level comparison to identify where the scaffolding works as intended and where redesign, potentially incorporating supplemental instruction, peer-mentoring, or flipped-classroom elements, might be required.
A recent meta-analysis reinforces the enduring significance of scaffolding for academic progression. Synthesizing 313 studies, [72] reported that regulated-learning scaffolds improved performance by an average effect size of 0.31, particularly in sequential STEM curricula where prior-knowledge gaps amplify failure risks. These results empirically support this thesis's modeling of prerequisite–core linkages as a determinant of student success.

2.4 Pandemic Disruption Beyond Delivery Mode

The pandemic was not just a platform change; it cut into continuity, routine, and motivation. K–12 meta-analyses and national datasets report significant learning losses, especially in mathematics, with uneven recovery across subjects and populations [87]. In higher education, qualitative and survey research from 2021 to 2025 shows a consistent pattern: institutions executed emergency remote teaching quickly, but students experienced variable quality, stress, and patchy support even when satisfaction with safety and flexibility went up [56], [80], [82]. Recent syntheses also note that many students and employers have re-tilted toward face-to-face or hybrid modes for collaboration-heavy learning after the pandemic peak [83].
For this research study, the timeline matters. In BRAC's modular semester calendar, COVID disruptions did not just pause face-to-face learning but reshuffled assessment schedules, altered RS planning, and temporarily relaxed progression rules. We explicitly compare pre-COVID, during-COVID, and post-COVID semesters using period indicators in models, checking whether any short-term boosts, such as grading leniency or policy accommodations, were followed by dips when in-person assessment norms returned.
Post-2023 scholarship has extended beyond delivery shifts to examine lingering pandemic effects on equity and well-being. [86] showed that inclusive residential programming in multicultural universities reduced GPA disparities by 0.2–0.4 points through immersive peer-support interventions, highlighting how COVID-19 exacerbated structural divides but also prompted corrective belonging-based frameworks. These findings emphasize the importance of holistic institutional analyses that integrate social and emotional dimensions alongside pedagogical recovery.

2.5 Class Size and Learning Outcomes

The narrative that small classes are better is well established in primary and secondary education research, but the higher education evidence is more nuanced and context-specific. A large-scale 2021 multi-institutional study found that an increase in class size was associated with lower average grades (roughly minus 0.08 standard deviations), but the magnitude varied substantially across disciplines and course types [52]. For example, seminar-based humanities classes appeared more sensitive to size increases than lecture-driven introductory science courses. Earlier quasi-experimental evidence also indicated that larger sections tend to depress student satisfaction and instructor evaluations, with some instructors compensating by simplifying assessments or reducing the frequency of feedback [22].

Emerging research suggests the relationship between class size and outcomes is non-linear, where student performance remains relatively stable up to a certain enrollment threshold, beyond which declines become more pronounced [74]. The threshold itself appears to be discipline-dependent, shaped by content density, reliance on discussion, and the need for individualized feedback. Instructional design plays a critical mediating role: courses built around structured collaboration, frequent low-stakes assessments, and active learning strategies have shown the ability to offset many of the disadvantages of larger class sizes [38], [47].

Equity considerations are increasingly important in this discussion. Studies have shown that active learning disproportionately benefits historically underrepresented or academically at-risk groups, thereby narrowing performance gaps [47]. This is particularly relevant for BRAC, where section sizes range from intimate RS cohorts of under 30 students to large main-campus lecture halls exceeding 80 to 100 students, creating natural variation in peer interaction and instructor contact. This means that in resource-constrained settings like Bangladesh, the class size problem may not be solved by simply lowering numbers, but by embedding interactive, inclusive teaching practices that leverage available human and technological resources. In this thesis, class size is measured at the precise section–term level, allowing for within-program analysis that avoids masking effects caused by cross-discipline variation.

Emerging work on intensive or immersive scheduling indirectly supports the benefits of smaller classes. [70] found that low-enrollment, compressed-format courses improved achievement by 8–12% in Australian higher education, an effect comparable to reduced class sizes enhancing feedback and engagement loops. Future research could quantify optimal section-size thresholds that yield measurable CGPA gains in resource-limited contexts such as Bangladesh.

2.6 Residential Semester (RS) as a Learning Environment

There is limited formal literature on BRAC University's RS, but there are close analogs. Living-learning communities and first-year learning communities consistently show improvements in retention, early-year GPA, and a stronger sense of belonging, including randomized and quasi-experimental evidence [58], [62], [75]. Studies from 2021 to 2024 also report that these communities mitigate equity gaps and help students navigate large campuses by anchoring them in smaller, high-contact cohorts [62]. In pharmacy and other applied fields, students reported a preference for hybrid designs post-COVID, that is, keep live sessions for contact, but provide recordings for spaced review and self-pacing [68].

RS shares some of these features: cohorting, fewer distractions, and dense scheduling. Unlike many learning community models in Western and local contexts, BRAC's RS is compulsory for all undergraduates and physically relocates students to a rural residential campus, altering both their study habits and social networks in ways not replicated on the main campus. This research tests whether those conditions translate into measurable GPA differences compared to main campus terms, and whether effects vary by department.

Recent investigations further validate the GPA-enhancing potential of immersive learning environments analogous to RS. [70] compared immersive semesters to conventional ones and found 10–15% higher achievement, largely due to increased instructor feedback in smaller cohorts. Extending this, [86] reported 0.3-point GPA uplifts in inclusive residential programs that fostered student belonging post-COVID. Similarly, [77] demonstrated that immersive simulation-based environments improved time-management behaviors and outcomes, moderated by baseline GPA. These findings align with the positive RS correlations observed in this study and underline the pedagogical value of immersion and community.

2.7 Failure Rates and Structural Problems

High D, F, and Withdrawal rates are red flags. In many systems, gateway courses carry higher DFW rates than smaller, upper-division courses and often show equity gaps [50], [57], [84]. Evidence from long-running redesigns, for example Peer-Led Team Learning in General Chemistry, shows that structured, active support can lift success and reduce DFW at scale over decades [81]. During COVID, some courses saw temporary dips in failure due to likely grading policy shifts or emergency leniency, followed by a correction when face-to-face norms returned. In South Asia, qualitative studies also linked performance swings to uneven access and mental-health stressors [80].

In this thesis, we map year-by-year failure trends and tie spikes or dips to policy or context changes, such as pandemic periods or assessment format shifts. At BRAC, such mapping also allows identification of whether RS terms systematically differ from main-campus terms in failure rates, and whether high-risk courses overlap with high-impact courses from the previous subsection. Courses crossing high-risk thresholds, for example sustained 20 to 25 percent failure, are flagged for syllabus pacing, formative assessment, and support redesign.

Recent post-pandemic analyses have begun pinpointing structural sources of STEM attrition. [66] found that prerequisite mismatches and overloaded syllabi accounted for 20–30% spikes in failure rates within gateway courses, recommending redesigned scaffolding and balanced assessment structures for equity. These patterns closely echo the high-failure dynamics identified in this dataset and reinforce the need for continuous curricular auditing.

2.8 Magnitude of Correlations in Educational Data and Methodological Safeguards

In educational research, observed associations between student metrics such as course grades, GPA, and academic outcomes commonly fall in the *moderate* range rather than approaching unity. Cohen's benchmarks ($|r| \approx 0.10$ = small, $|r| \approx 0.30$ = medium, $|r| \approx 0.50$ = large) are often used heuristically in social science domains and are considered realistic in the presence of measurement error and heterogeneity of constructs [7]. Indeed, educational datasets rarely exhibit correlations beyond

0.60 unless there is structural overlap or shared variance due to compositional artifacts. i.e. the two things being compared are closely related or even partially made up of the same information [34] [63].
A central constraint on observed correlation magnitudes is **attenuation due to measurement error**: imperfect reliability in either variable reduces the maximal observable r (i.e., the *attenuation ceiling*). Classic psychometric theory shows that the "true" correlation is bounded by

$$r_{\text{observed}} \leq \sqrt{\rho_x \, \rho_y}$$

where ρ_x and ρ_y are the reliabilities of the two variables (e.g., grades, GPA) [21] [63]. Because academic grades and GPA composites include error and variance across courses, even strong underlying associations often manifest as moderate observed r values.
Another methodological pitfall arises when correlating a focal course grade with a cumulative GPA that itself *includes* that grade. This **part–whole inflation** artificially increases the shared variance, inflating correlation estimates. To avoid such bias, psychometric and educational methodology texts recommend excluding the focal item from the composite (i.e., a "leave-one-out" approach) or using alternate composite constructions [4] [23].
When researchers run multiple pairwise tests (e.g., many courses correlated with CGPA), controlling for **multiple comparisons** is crucial. The Benjamini–Hochberg procedure, which controls the False Discovery Rate (FDR), offers greater power than Bonferroni-style controls while maintaining acceptable error bounds [11] [14]. Coupled with effect-size thresholds and reporting of confidence intervals, this approach is now widely recommended in multi-test settings [11].
Overall, observed **r** values in the range of 0.30–0.60 in educational settings are both common and meaningful once measurement imperfections and appropriate methodological safeguards are considered. Values approaching 1.0 in this context typically suggest redundancy, mechanical overlap, or circularity rather than genuinely stronger substantive relationships.
In this study, therefore, we adopt a stringent significance level of $p < 0.001$, ensuring that detected relationships reflect substantial and replicable effects rather than statistical artifacts of large samples.

2.9 Synthesis

Delivery mode, timing, class size, scaffolding, and micro-environment all interact. A one-size prescription will not hold across programs. The literature supports a targeted approach by identifying leverage courses, rebuilding large classes around active methods, tailoring support by program, and keeping hybrid affordances where they clearly help. Few prior studies in the Global South have combined these structural factors into a single, multi-model framework; this thesis does so while retaining program-level granularity, allowing patterns to emerge that would be hidden in aggregated institutional data. The analyses in this thesis are program-aware, time-aware, and course-aware, and the modelling choices follow from these design questions.

Recent advances in educational-data-mining (EDM) methodologies emphasize hybrid statistical–machine-learning models for interpretability and accuracy. Meta-analyses such as [72] validate the combined use of regression and Generalized Estimating Equations (GEE) when modeling scaffolded and longitudinal academic effects, supporting their selection in this study over purely algorithmic ML approaches.

2.10 Statistical Modelling: What Is Used Here and Why

Correlation and impact scoring. Correlations quantify course–CGPA links; multiplying by log enrollment prioritizes courses with both influence and reach. This is a screening device, not a causal claim. It flags where we should invest attention first [18] [31].

Multiple linear regression. Multivariate regression tests whether a small set of high-leverage course grades jointly predict CGPA after controlling for others, and surfaces multicollinearity or redundancy. Standard texts remain the reference for diagnostics, such as residuals, influence, and variance inflation factors, and validation [27]. Interaction terms, for example course by program, are probed and interpreted using established procedures [10].

ANOVA and GEE for periods and modes. One-way or two-way ANOVA compare mean GPA across delivery modes and COVID periods. For repeated observations, that is longitudinal students, Generalized Estimating Equations model within-student correlation without heavy distributional assumptions, using working correlation structures such as exchangeable and robust sandwich standard errors [5] [29] [24].

Non-parametrics when needed. When normality or equal variance assumptions wobble, Welch's t-test, Mann–Whitney U, and Kruskal–Wallis are used to cross-check results. These are standard tools for unequal variances and skewed distributions [2] [1] [3] [64].

2.11 Conceptual Mapping - TOE/HOT Frameworks

Scholarly investigations into student performance increasingly emphasize the interplay between technological infrastructure, organizational processes, and human adaptability. The *Technology–Organization–Environment (TOE)* framework introduced by [9] provides a robust foundation for analyzing how institutional and contextual factors jointly influence adoption, innovation, and performance outcomes. Originally developed for technological innovation studies, TOE has since been extended to higher education to examine e-learning readiness, digital transformation, and institutional resilience [35] [32].

Complementing this, the *Human–Organization–Technology (HOT-fit)* model by [19] offers a performance-oriented lens that integrates system quality, user satisfaction, and organizational support. Derived from the DeLone and McLean IS Success Model [13], HOT-fit highlights the alignment between human competencies, technological adequacy, and institutional context as key determinants of effectiveness. Recent educational studies have used the HOT-fit model to evaluate learning management

systems, digital pedagogy, and user acceptance in blended and online learning environments [30] [43].
Building on these theoretical precedents, this study synthesizes elements from both frameworks to conceptualize **Academic Performance Outcomes (APO)** as a multi-dimensional construct shaped by human, technological, organizational, and environmental dynamics. This integrative perspective bridges individual-level behavior with institutional structure, providing a coherent analytical basis for examining how universities can design resilient, equitable, and technology-enabled learning ecosystems.

Chapter 3

Methodology

This chapter presents the methodological framework adopted in the study, beginning with the systematic collection of primary academic performance data from institutional sources. It then describes the sequential processes of data preprocessing, feature engineering, and exploratory data analysis undertaken to prepare the dataset for robust statistical examination. The chapter further details the analytical models employed to address each research hypothesis, including the rationale for model selection, procedures for testing statistical assumptions, and the specific implementation steps. This structured approach was designed to ensure methodological transparency, facilitate reproducibility, and maintain alignment with established statistical standards and the university's data governance policies.

3.1 Conceptual Framework

Understanding student performance in higher education requires an integrative perspective that captures both human and institutional dynamics. This study therefore adopts a multi-dimensional conceptual framework grounded in the principles of the *Technology–Organization–Environment (TOE)* framework proposed by [9] and the *Human–Organization–Technology (HOT-fit)* model developed by [19]. Both frameworks have been widely applied across disciplines to analyze technology adoption, performance outcomes, and system-level effectiveness in educational and organizational contexts [13] [35].

In line with these theoretical precedents, the present study extends these frameworks to model **Academic Performance Outcomes (APO)** as the result of four interrelated dimensions—**Human**, **Technological**, **Organizational**, and **Environmental**. Each dimension represents a structural or contextual influence on institutional learning outcomes, aligning with the study's hypotheses and subsequent data-driven analyses.

- **Human Dimension:** Captures student readiness, prior academic performance, faculty adaptability, digital competency, and program or major track.
- **Technological Dimension:** Reflects factors such as delivery mode (online, offline, or blended), IT infrastructure, platform integration, and changes in assessment format during the pandemic.

- **Organizational Dimension:** Relates to departmental policies, administrative and resource planning, and institutional management support that shape learning and evaluation environments.

- **Environmental Dimension:** Incorporates broader contextual factors including the pandemic period, campus setting (Residential vs. Main Campus), and cohort size or term-level effects.

Collectively, these dimensions form an integrated model that links institutional structure and external context to academic performance and resilience (demonstrated in Figure 3.1). This conceptual framework provides a theoretical foundation for the empirical analyses that follow and supports the study's hypothesis-driven exploration of structural determinants of student outcomes.

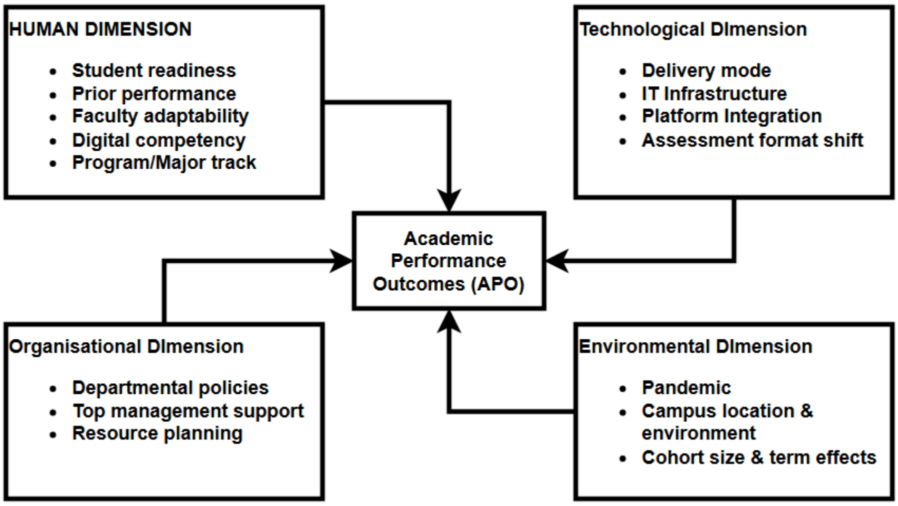

Figure 3.1: Conceptual framework adapted from the Technology-Organization-Environment (TOE) framework [9] and the Human-Organization-Technology (HOT-fit) model [19]. The model integrates four dimensions to explain structural and contextual influences on Academic Performance Outcomes (APO).

3.2 Data Collection and Scope

The dataset comprises primary, first-party data obtained directly from BRAC University's academic management systems through authorized institutional credentials. Records were retrieved at the granularity of student × course × semester × section, allowing detailed tracking of academic performance over time and across contexts. Our operationalization of student success indicators and curriculum structures aligns with established higher-education effectiveness literature [31].
The full dataset spans two decades of academic history, from **Spring 2004** to **Fall 2023**, and covers multiple departments, programs, and delivery modes. After preprocessing, several hundred thousand valid records remained for analysis. Core fields included:

- Student ID (anonymized unique identifier) and Student Name (removed post-linkage)

- Marks obtained and letter grades with standardized GPA conversion
- Course code, section, credit hours, course type (core, elective, lab, theory)
- Semester and year identifiers
- Faculty assigned (used only for exploratory checks)
- Campus location tags (Main Campus or Residential Semester, hereafter RS)
- Each record was labeled as Offline, Online, or Hybrid from scheduling sources; transitional "hybrid-like" emergency terms during COVID were coded under Hybrid when synchronous+asynchronous elements co-existed.

The scope and composition are summarized in Table 3.1. The diversity of time periods, departments, and delivery modes enables robust comparisons across institutional and environmental contexts.

Table 3.1: Summary of Dataset Characteristics

Attribute	Value
Years Covered	2004–2023
Total Records (Rows)	890,998
Unique Students	37,954
Unique Courses	886
Unique Semesters	50
Offline Delivery Mode Records	430,155
Online Delivery Mode Records	199,495
Hybrid Delivery Mode Records	261,348
Main Campus Records	822,495
Residential Semester (RS) Records	68,503
Number of Departments/Programs	17

3.3 Data Processing Pipeline

Transforming raw academic records into a reliable analytical dataset required a rigorous, multi-stage data processing pipeline. Given the large scale and longitudinal nature of the institutional dataset, the pre-processing phase was designed both to clean and standardize information as well as to preserve the contextual integrity of each record. The goal was to construct a unified dataset that accurately reflects students' academic trajectories while remaining analytically compatible with statistical and regression models used later in this study.

The pipeline integrated both automated and manual procedures to address missing data, multiple course attempts, inconsistent naming conventions, and mixed-format semester identifiers. Each stage of the process—ranging from raw extraction to engineered features—was governed by reproducible, script-based logic to minimize human error and maintain transparency in data handling. The final dataset thus represents a refined and structured foundation for the empirical analyses presented in subsequent chapters.

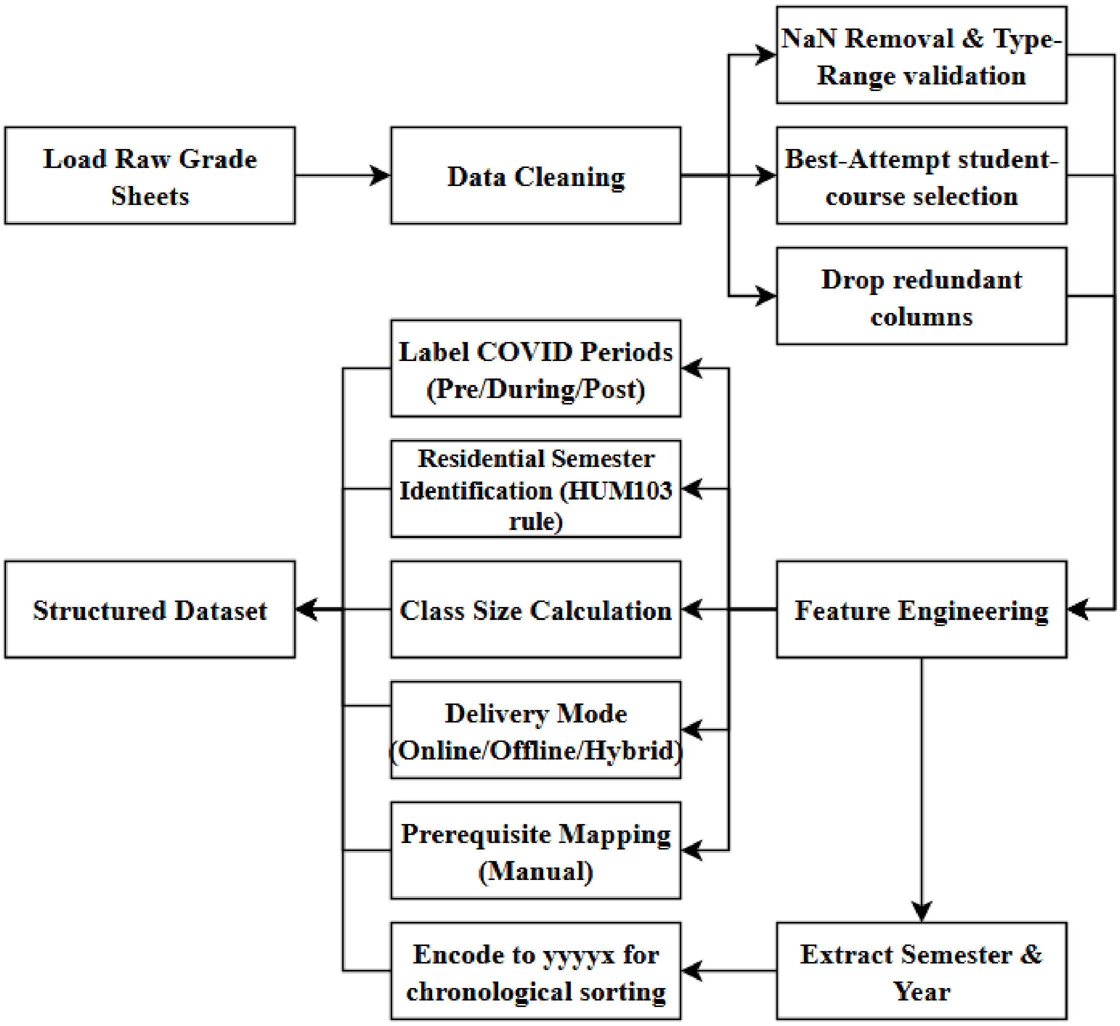

Figure 3.2: Data Processing Workflow from raw grade sheets to a structured analysis dataset. The pipeline covers NaN removal, best attempt selection, and redundant column removal; semester and year extraction; feature engineering including encoding to yyyyx for chronological sorting, manual prerequisite mapping, delivery mode tagging, class size calculation, RS identification, and COVID period labeling.

The main steps of the workflow depicted by Figure 3.2 is given below:

3.3.1 Standardization and De-duplication

To ensure consistency and analytical reliability, the dataset went through a structured standardization process prior to statistical testing. This step aligned the course identifiers, resolved duplicate entries, and filtered out incomplete or non-academic records. Special care was taken to distinguish between repeated course attempts and valid single attempts, as this distinction directly influences the accuracy of GPA-related analyses. The following procedures summarize the key normalization and de-duplication steps applied:

- **Normalization of identifiers:** Course codes were reformatted to a consistent scheme (e.g., `CSE220`).
- **Missing value removal:** Records with missing GPA, marks, or CGPA were excluded.
- **Attempt handling (analysis-specific):**
  - For CGPA-linked and course–CGPA impact analyses (H2), the *best transcript attempt* per course was retained to reflect official standing.
  - For failure-rate analyses (H7) and prerequisite→core analyses (H3), *first attempts* were used to avoid upward bias from repeats.
- **Filtering:** Internships, theses, and administrative repeats were excluded.
- **Sorting and indexing:** Data were ordered by `student_id`, course, and `year_semester`.

3.3.2 Derived Tags and Enrichment

- **Semester and year encoding:** From raw labels, separate semester and year variables were extracted, then combined into a sortable code yyyy$x$, where Spring = 1, Summer = 2, Fall = 3.
- **COVID period classification:** Pre-COVID up to Fall 2019 (20193), During-COVID from Spring 2020 to Summer 2021 (20201–20212), Post-COVID from Fall 2021 (20213) onward, based on official calendars.
- **Residential Semester identification:** A semester was flagged RS if it contained at least one RS-exclusive course (EMB101, DEV101, HUM103, and in recent years BNG103). Otherwise, it was labeled Main Campus. Labels were cross-verified against academic records.
- **Class size computation:** Students per course_code × section × year_semester were counted for class size analyses.
- **Delivery mode tagging:** Each record was labeled as Offline, Online, or Hybrid from scheduling sources, consistent with the literature on online course design and performance [38].
- **Prerequisite mapping:** Where applicable, prerequisite to core links were mapped to enable course-pair analyses.

The pipeline was implemented as reproducible scripts to minimize manual handling and ensure consistent runs across analyses.

3.4 Exploratory Data Analysis

Prior to formal testing, an exploratory analysis combined descriptive statistics and visualization to assess distributions, outliers, and initial patterns. Violin plots were used for GPA distributions across COVID periods and delivery modes, boxplots for variation across programs and semesters, line plots for temporal trends, and bar plots for mean GPA by key categorical factors. Observed deviations from normality motivated the inclusion of non-parametric tests that complement parametric models, ensuring valid conclusions when assumptions are only partially met [12]. The visualizations and analysis was done in Chapter for as required per hypothesis.

3.5 Feature Engineering

After data cleaning and validation, several derived variables were systematically engineered to capture complex academic dynamics not directly observable in the raw dataset. Feature engineering helped with translating institutional records into analytically meaningful indicators that could support hypothesis-specific testing. These transformations enabled deeper insight into structural relationships, such as the impact of individual courses, the strength of prerequisite linkages, and the consistency of performance patterns across programs and semesters.
By designing targeted variables, the dataset evolved beyond a static grade archive into a framework capable of revealing statistical dependencies and performance trends within the university's academic ecosystem. The most relevant engineered features are summarized below:

- **Impact score:** Combines Pearson correlation between course GPA and CGPA with course frequency to emphasize widely taken courses.
- **Performance deltas:** Differences between prerequisite and subsequent core course GPAs.
- **GPA variance and dispersion:** Course-level spread indicators for stability checks across programs and time.

3.6 Research Hypotheses

The study evaluates nine hypotheses aligned with the institutional context and the literature:

H1 Differences in GPA by delivery mode (Online, Offline, Hybrid).

H2 Identification of individual courses with strongest association with CGPA (high-impact courses).

H3 Predictive influence of prerequisite course GPA on corresponding core course GPA.

H4 Departmental variation in the academic impact of the COVID-19 pandemic across disciplines and course types.

H5 Temporal shifts in GPA across pandemic phases (Pre-COVID, During-COVID, Post-COVID), reflecting adaptation and recovery patterns.

H6 Relationship between class size and GPA across disciplines.

H7 Residential Semester (RS) vs. Main Campus performance differences.

H8 Identification of courses with consistently high failure/withdrawal rates and their structural patterns.

H9 Program- and department-level performance differences derived from H1-H8.

3.7 Variable Definitions

A clear definition of variables is essential to ensure analytical transparency and reproducibility. Each variable in the dataset was systematically categorized based on its analytical role—whether as an outcome measure, explanatory factor, or derived construct. This classification guided model selection and hypothesis mapping by clarifying how each feature contributes to understanding academic performance dynamics.

- **Dependent variables:** Course GPA, cumulative GPA.
- **Independent variables:** Course ID, course type, frequency, class size, semester code, campus, delivery mode, department or program.
- **Derived variables:** Impact scores, mean GPA per course, GPA variance, prerequisite to core performance difference, COVID period, RS flag.

3.8 Statistical Significance and Reliability Controls

Given the large scale of the dataset and the multiplicity of pairwise tests, this study adopts conservative statistical criteria to ensure that identified relationships are both reliable and substantively meaningful.

3.8.1 Statistical Significance, Multiplicity, and Effect-Size Criteria

For each candidate association (e.g., course grade $\rightarrow$ CGPA), we compute the Pearson correlation coefficient r (two-sided) at a nominal significance level of $\alpha = 0.001$. Because our analysis involves multiple simultaneous tests, we apply the **Benjamini–Hochberg procedure** to control the false discovery rate (FDR) at $q = 0.05$ [11]. Only associations that satisfy both $p < 0.001$ and pass the FDR adjustment are retained in our primary reporting. To further guard against detection of trivial associations in large samples, we impose a minimum effect-size threshold of $|r| \geq 0.30$. Whenever relevant, we also compute and report 95% confidence intervals for r using Fisher's z-transformation.

3.8.2 Mitigating Part–Whole Bias via LOCO-CGPA

To avoid inflation of correlation coefficients due to inclusion of the focal course grade within the CGPA composite, we derive a **leave-one-course-out CGPA (LOCO-CGPA)** for each test. That is, when assessing course X vs. CGPA, we compute CGPA without including the grade from X. All correlations and regressions reported with CGPA refer to this LOCO construct. This step minimizes mechanical overlap and yields more conservative, interpretable estimates [4] [23].

3.8.3 Reliability Considerations and Interpretation

Observed correlations are interpreted in light of theoretical attenuation: the measured r typically underestimates the "true" population correlation due to measurement error in both variables. While no formal correction-for-attenuation is applied in the main analysis, we note that moderate observed r values (e.g., 0.30–0.60) can correspond to stronger underlying relationships when variable reliability is high [21] [63]. This interpretive framework aligns with established benchmarks for educational data [7] [34].

3.8.4 Complementary Predictive Modeling

Beyond bivariate associations, we fit program-level multivariate linear models using the top correlated courses as predictors of CGPA, and evaluate **cross-validated** R^2 (CV-R^2) to assess out-of-sample explanatory power. This ensures that results emphasize predictive strength rather than solely in-sample fit, reinforcing the robustness of identified relationships.

3.9 Statistical Models and Analytical Procedures

Model selection was guided by the structure of the data, the nature of variables, and the plausibility of standard assumptions. Assumption checks included normality of residuals, homogeneity of variances, multicollinearity diagnostics, and independence. Non-parametric alternatives were employed when assumptions were not satisfied.

3.9.1 Correlation and Regression

To examine relationships among academic performance variables, both correlation and regression analyses were employed. These techniques provided complementary perspectives: correlation quantified the strength and direction of linear associations, while regression models estimated the magnitude and significance of those effects after accounting for multiple influencing factors. Together, they formed the analytical backbone for testing hypotheses related to course-level impact, prerequisite dependencies, and broader structural determinants of GPA variation.

(a) Pearson Correlation

The **Pearson product–moment correlation coefficient** is a foundational statistical measure used to quantify the **strength and direction of a linear relationship** between two continuous variables [27] [28]. In this study, it was employed to

examine associations such as the relationship between individual course GPA and overall cumulative GPA (CGPA).
Mathematically, the coefficient is expressed as:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{3.1}$$

where X_i and Y_i represent **paired observations** (e.g., GPA in a specific course and cumulative GPA), while $\bar{X}$ and $\bar{Y}$ denote their respective **means**. The numerator captures the **covariance** between the two variables—indicating how they co-vary—while the denominator standardizes this value by the product of their **standard deviations**, thereby constraining r to a range between -1 and $+1$.
A **positive correlation** ($r > 0$) indicates that higher values of one variable tend to coincide with higher values of the other (e.g., good performance in a key course aligning with higher CGPA), whereas a **negative correlation** ($r < 0$) suggests an inverse relationship. An r value near zero implies **little to no linear association**. However, Pearson correlation measures only **linear dependence**; it may underestimate relationships that are nonlinear or influenced by **outliers**. To mitigate this, diagnostic checks were performed to ensure approximate linearity and to assess the influence of extreme values before interpretation.
In this research, the Pearson coefficient informed both the **course impact analysis** (H2) and the **prerequisite-to-core course linkage analysis** (H3), serving as a quantitative indicator of how strongly performance in one context predicts outcomes in another.

(b) Linear and Multiple Regression

Linear regression is a fundamental statistical modeling technique used to examine how one continuous variable (**dependent variable**) changes as a function of another (**independent variable**). In its simplest form, it assumes a straight-line relationship between the predictor and the outcome [16], [27], [28]. The general equation is expressed as:

$$Y = \beta_0 + \beta_1 X + \epsilon, \tag{3.2}$$

where Y represents the dependent variable (e.g., cumulative GPA), X denotes the independent variable (e.g., course GPA), β_0 is the **intercept** or baseline value of Y when $X = 0$, β_1 is the **slope coefficient** indicating the expected change in Y for each one-unit change in X, and ϵ represents the **random error term** capturing unexplained variation.
Multiple regression extends this framework by incorporating two or more predictors simultaneously:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon, \tag{3.3}$$

allowing the model to estimate the **unique contribution** of each predictor while controlling for the effects of others. This makes multiple regression particularly suitable for complex educational datasets, where factors such as course type, class size, or delivery mode jointly influence academic outcomes.
Model evaluation involved several diagnostic measures. The **coefficient of determination** (R^2) was used to quantify how much variance in the dependent variable is

explained by the predictors, while **residual analysis** verified assumptions of linearity, normality, and homoscedasticity. Additionally, **Variance Inflation Factors (VIF)** were computed to detect multicollinearity among predictors, ensuring the stability of coefficient estimates.
In this study, regression models supported hypothesis testing across multiple dimensions: identifying high-impact courses (H2), assessing the predictive strength of prerequisites (H3), and quantifying contextual factors such as delivery mode and class size (H1, H5). Together, these models provided a statistically grounded understanding of how individual and structural variables shape student performance.

3.9.2 Group Comparisons

To evaluate whether structural and contextual factors produced statistically significant differences in academic outcomes, several group comparison tests were employed. These methods enabled the study to contrast mean or median GPA values across distinct categories such as delivery mode (online, in-person, hybrid), learning environment (Residential Semester vs. Main Campus), and academic period (pre-, during-, and post-COVID). Each test was selected based on underlying data characteristics, such as distributional assumptions and variance equality, ensuring methodological appropriateness and robustness of inference.
By combining both **parametric** and **non-parametric** approaches, the analysis accommodated variations in sample size, heterogeneity of variance, and non-normal distributions common in large educational datasets. Collectively, these tests provided a rigorous foundation for examining whether observed performance differences among groups were statistically meaningful or attributable to random variation.

(a) Welch's t-test

The **Welch's t-test** was employed to compare the means of two independent groups when the assumption of equal variances was not met. Unlike the traditional Student's t-test, which assumes homogeneity of variance, Welch's formulation adjusts both the test statistic and its degrees of freedom to account for unequal sample sizes and variance heterogeneity [12]. This makes it particularly suitable for institutional data such as that used in this study, where cohorts (e.g., Residential Semester vs. Main Campus) often differ in enrollment scale, grading dispersion, and contextual conditions.
In mathematical terms, the test evaluates whether the mean difference between two groups is statistically significant beyond what might be expected by random variation. The test statistic is computed as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{3.4}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means, s_1^2 and s_2^2 are the sample variances, and n_1 and n_2 denote the sample sizes of the two groups. The resulting t value is evaluated against a reference t-distribution with degrees of freedom adjusted using the Welch–Satterthwaite equation.
Within the context of this research, the Welch's t-test was primarily applied to compare mean GPA values between **Residential Semester (RS)** and **Main Campus**

students, capturing the influence of differences in location, living environment, and academic structure. It also supported subgroup comparisons, such as between **online and in-person** learning modes during the pandemic period. These analyses directly extend the correlation and regression findings discussed in Section 3.9.1, offering inferential validation for observed differences in academic performance patterns.

(b) Mann–Whitney U Test

The **Mann–Whitney U test** (also known as the Wilcoxon rank-sum test) is a **non-parametric alternative** to the independent samples t-test, suitable when the assumption of normality is violated or when data are ordinal rather than continuous [8], [12]. Instead of comparing group means, the test examines whether one group tends to yield systematically higher or lower values than another by analyzing the **rank ordering** of observations across both samples.

The test statistic is calculated based on the sum of ranks assigned after combining observations from both groups. It effectively tests whether the probability that a randomly selected value from one group exceeds a randomly selected value from the other differs significantly from 0.5. The U statistic is expressed as:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1, \tag{3.5}$$

where n_1 and n_2 represent the sample sizes of the two groups, and R_1 is the sum of ranks for the first group. The resulting U value can be converted into a z-score approximation for large samples, enabling inference using the standard normal distribution.

In this study, the Mann–Whitney U test was applied to evaluate GPA differences between student cohorts when data distributions deviated from normality—particularly for comparisons between **Residential Semester (RS)** and **Main Campus** environments or between **online and in-person** delivery modes during the pandemic. This approach complemented the parametric Welch's t-test (Section 3.9.2) by providing a distribution-free method to confirm whether observed performance differences were statistically significant under non-Gaussian conditions.

(c) ANOVA

The **Analysis of Variance (ANOVA)** is a parametric statistical method used to determine whether there are significant mean differences among three or more independent groups [15], [27], [28]. It extends the logic of the t-test by partitioning the total variation in a dataset into components attributable to **between-group** and **within-group** sources. The ratio of these variances forms the **F statistic**, calculated as:

$$F = \frac{\text{Between-group Variance}}{\text{Within-group Variance}}. \tag{3.6}$$

A higher F value indicates that the variability between group means is substantially greater than the variability within groups, suggesting statistically significant differences. One-way ANOVA examines a single categorical factor, while two-way ANOVA extends the analysis to two factors simultaneously, enabling the assessment of both main and **interaction effects**.

In this study, **one-way ANOVA** was used to test mean GPA differences across categorical factors such as **COVID periods**, **delivery modes**, and **campus types**, while **two-way ANOVA** was employed when interactions—such as between **delivery mode and department**—were theoretically plausible. These analyses provided a comprehensive understanding of how structural and contextual dimensions jointly influenced academic outcomes.

Where assumptions of normality or homogeneity of variance were violated, the analysis reverted to **non-parametric alternatives** such as the Kruskal–Wallis H-test (Section 3.9.2) to maintain inferential validity. This hybrid approach ensured robustness while adhering to statistical best practices for educational data analysis.

(d) Kruskal–Wallis H-test

The **Kruskal–Wallis H-test** is a **non-parametric alternative** to the one-way ANOVA, used to evaluate whether three or more independent groups originate from the same population distribution [3], [12]. Unlike ANOVA, which relies on assumptions of normality and homogeneity of variance, the Kruskal–Wallis test operates on the **ranks** of the data rather than the raw values. This makes it particularly robust when GPA distributions are skewed, contain outliers, or exhibit unequal variances—conditions frequently encountered in large-scale academic datasets.

The test statistic, denoted by H, is computed as:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \bar{R}\right)^2, \tag{3.7}$$

where k is the number of groups, n_i is the sample size of group i, $\bar{R}_i$ is the average rank of observations within that group, $\bar{R}$ is the overall mean rank, and N is the total number of observations across all groups. Under the null hypothesis—that all groups have identical population medians—the statistic approximately follows a chi-square (χ^2) distribution with $(k-1)$ degrees of freedom.

In this study, the Kruskal–Wallis H-test was applied in scenarios where the assumptions of ANOVA were violated, such as comparing GPA distributions across **multiple departments**, **COVID periods**, or **delivery modes** with evident non-normality. When significant differences were detected, **post-hoc pairwise comparisons** using adjusted rank-based methods were conducted to identify which specific groups differed. By complementing parametric tests like ANOVA (Section 3.9.2), the Kruskal–Wallis test strengthened the robustness and interpretability of the overall inferential framework, ensuring that results remained valid even under heterogeneous and non-Gaussian data conditions.

3.10 Dependence and Distribution Tests

Understanding the relationships between categorical academic variables—such as grade distributions, delivery modes, and departmental classifications—required statistical methods that could test for **association rather than numerical difference**. While correlation and regression (Section 3.9.1) focus on continuous variables, dependence tests like the Chi-Square Test provide a framework for evaluating whether categorical outcomes differ systematically across groups. These tests were particularly relevant for examining patterns such as **failure-rate disparities**, **grade distribution asymmetries**, and **mode-specific performance variations** within the institutional dataset.

3.10.1 Chi-Square Test for Independence

The **Chi-Square Test for Independence** is a non-parametric statistical procedure used to determine whether two categorical variables are statistically independent or associated [12], [17]. It compares the observed frequencies in a contingency table against the frequencies expected under the assumption of independence. The test statistic is expressed as:

$$\chi^2 = \sum \frac{(O - E)^2}{E}, \tag{3.8}$$

where O represents the **observed frequencies** and E the **expected frequencies** for each cell, calculated as $E = \frac{\text{(row total)}\times\text{(column total)}}{\text{grand total}}$.
The calculated χ^2 statistic follows a chi-square distribution with $(r-1)(c-1)$ degrees of freedom, where r and c denote the number of rows and columns in the contingency table, respectively. A significant χ^2 value indicates that the distribution of one categorical variable differs systematically across the levels of another, suggesting dependence between them.
In this study, the Chi-Square Test for Independence was applied to assess whether grade distributions, failure rates, or pass/fail outcomes were associated with factors such as **delivery mode (online vs. in-person)**, **campus type (RS vs. Main)**, or **departmental affiliation**. Expected count checks were performed to ensure that at least 80% of cells had expected frequencies greater than five, satisfying the test's validity conditions [28]. In cases where this assumption was violated, categories were merged or alternative exact tests were considered.
The results from this analysis complemented findings from ANOVA and Kruskal–Wallis tests by extending inference from continuous GPA measures to categorical grade-level outcomes. This combination of parametric, non-parametric, and categorical methods provided a comprehensive and multidimensional understanding of academic performance structures within the university dataset.

3.11 Longitudinal and Repeated Measures

To accommodate repeated measures at the student level, GEE with an exchangeable correlation structure estimated population-averaged effects across time and contexts [25]. This addresses within-student correlation without specifying subject-specific random effects.

3.11.1 Generalized Estimating Equations (GEE)

The **Generalized Estimating Equations (GEE)** framework extends the generalized linear model (GLM) to accommodate **correlated observations** that arise from repeated measurements on the same subjects over time [6], [25]. In this study, each student contributed multiple GPA observations across semesters, creating intra-subject correlations that violated the independence assumption of traditional re-

gression models. GEE provides a flexible, semi-parametric solution by estimating **population-averaged effects** rather than subject-specific ones.
The basic form of the GEE is expressed as:

$$g(\mu_{it}) = \mathbf{X}_{it}\boldsymbol{\beta}, \tag{3.9}$$

where $g(\cdot)$ is a **link function** (e.g., identity or logit), $\mu_{it} = E(Y_{it})$ is the expected value of the response variable (e.g., GPA) for individual i at time t, $\mathbf{X}_{it}$ is the vector of covariates, and $\boldsymbol{\beta}$ represents the estimated regression coefficients. The covariance structure of the repeated observations is modeled through a **working correlation matrix** $R(\alpha)$, where α represents the correlation parameters.
In this research, an **exchangeable correlation structure** was specified, assuming that the correlations between all pairs of observations within a student are equal:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix}.$$

This choice reflects the reasonable assumption that a student's GPA performance across semesters is consistently correlated, regardless of the specific time gap between semesters.
Parameter estimates are obtained by solving the estimating equations:

$$\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{3.10}$$

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$ is the derivative matrix of the mean with respect to parameters, and $\mathbf{V}_i$ is the working covariance matrix of $\mathbf{Y}_i$. The use of a "working" covariance allows GEE to produce **robust (sandwich) standard errors** even when the specified correlation structure is imperfect.
In the context of this study, GEE was applied to estimate **population-level trends** in academic performance across delivery modes, semesters, and campus types, adjusting for repeated GPA measures within each student. This approach enabled the analysis to capture generalizable institutional effects (e.g., average impact of Residential Semester participation or online delivery) without requiring explicit modeling of individual-level random effects.

3.11.2 Mixed-Effects Models (LMM/GLMM)

In mathematical form, a basic linear mixed-effects model can be represented as:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \epsilon_{it}, \tag{3.11}$$

where Y_{it} denotes the observed outcome (e.g., GPA) for student i at time t; $\mathbf{X}_{it}$ is the design matrix for fixed effects with coefficients $\boldsymbol{\beta}$; $\mathbf{Z}_i$ is the design matrix for random effects with coefficients $\mathbf{b}_i$; and ϵ_{it} is the residual error term assumed to be normally distributed with mean 0 and variance σ^2.

The random effects $\mathbf{b}_i$ are typically assumed to follow a multivariate normal distribution:

$$\mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}), \tag{3.12}$$

where $\mathbf{D}$ is the covariance matrix of random effects. This formulation enables each student to have a unique intercept (and potentially slope), reflecting individual variation in baseline performance and trajectory over time.

For outcomes that were not strictly continuous or normally distributed (e.g., binary pass/fail or categorical outcomes), the framework was extended to a **Generalized Linear Mixed Model (GLMM)** using appropriate link functions such as the *logit* or *probit*. These models jointly account for non-Gaussian distributions and within-subject correlations through the random effects structure.

In this study, **student-level random intercepts** were specified to model repeated GPA observations within individuals, addressing within-student dependence while estimating **fixed effects** for factors such as **delivery mode**, **COVID period**, **department**, and their interactions. This structure was particularly important for testing key hypotheses—H1 (delivery mode effects), H4, H5 (COVID-period variation), and H7 (RS vs. Main Campus comparison)—where temporal and contextual influences intersected.

Model diagnostics included checks for convergence stability, examination of random-effects variance components, and inspection of residual patterns to confirm model adequacy. Together, the mixed-effects framework complemented the population-level approach of GEE by offering **subject-specific insights**, thus enabling a richer interpretation of both individual and institutional patterns in student academic performance.

3.12 Interactions and Moderation

Complex educational phenomena often emerge from the interplay of multiple factors rather than from isolated main effects. To capture these nuanced relationships, **interaction terms** were incorporated into regression and ANOVA frameworks to examine how the influence of one predictor variable changes depending on the level of another—an analytical approach known as **moderation analysis** [40], [49]. This method enables a deeper understanding of contextual dynamics such as whether the impact of **delivery mode** differs by **program**, **department**, or **COVID period**.

In a linear regression context, moderation is modeled through the inclusion of a multiplicative term between two predictors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \epsilon, \tag{3.13}$$

where Y is the outcome variable (e.g., GPA), X_1 and X_2 are predictor variables (e.g., delivery mode and program), and the coefficient β_3 quantifies the **interaction effect**. A statistically significant β_3 indicates that the relationship between X_1 and Y depends on the value or category of X_2. In ANOVA models, this principle extends naturally to test whether differences in group means depend jointly on two categorical factors (e.g., campus type and delivery mode).

Following the detection of significant interactions, **simple-effects analyses** were conducted to interpret the direction and magnitude of the moderation. These analyses decomposed the overall model into conditional relationships, clarifying under which conditions the moderating variable strengthened, weakened, or reversed the main effect. Results were visualized using **interaction plots** that displayed fitted GPA means or regression slopes across factor combinations, enhancing interpretability and transparency of effects [42], [45].

3.13 Model Limitations and Alternatives

Robust empirical analysis requires not only the application of appropriate statistical models but also a critical acknowledgment of their underlying assumptions, constraints, and methodological boundaries. In this study, multiple analytical frameworks—ranging from correlation and regression to ANOVA, non-parametric tests, and mixed-effects models—were employed to examine structural and contextual determinants of student performance. However, each model carries interpretive and technical limitations that influence the generalizability of results.

This section outlines, for each hypothesis, the primary analytical methods selected, the potential alternative approaches that were considered, and the reasons for inclusion or exclusion. It also highlights specific modeling limitations—such as sensitivity to variance heterogeneity, dependence on correct correlation structures, and convergence challenges in mixed-effects frameworks—that may affect the precision of statistical inference. By systematically articulating these methodological boundaries, the analysis maintains transparency and encourages cautious interpretation of results, while providing a rationale for the chosen balance between complexity, interpretability, and robustness.

3.13.1 Delivery Mode and COVID Period Analyses (H1, H4, H5; overlaps with H7)

The hypotheses addressing **delivery mode** (online, in-person, and hybrid) and **COVID-period effects** were among the most comprehensive in scope, encompassing both structural and temporal dimensions of student performance. Given that these factors intersected with campus type (Residential Semester vs. Main Campus) and extended across multiple semesters, analyses for H1, H4, H5 and partially H7 required models capable of handling both cross-sectional and longitudinal variation.

The **primary analytical framework** consisted of a combination of parametric, non-parametric, and repeated-measures techniques to ensure robustness under diverse data conditions. Specifically:

- **Two-way ANOVA** (Section 3.9.2) was employed to evaluate main and interaction effects between delivery mode and COVID-period classifications, allowing for the examination of whether GPA means differed significantly both across and within these factors.
- **Mixed-Effects Models (LMM/GLMM)** (Section 3.11.2) incorporated student-level random intercepts to account for repeated observations over time, modeling within-student correlations while estimating fixed effects for delivery mode, period, and their interactions. This approach captured individual heterogeneity and provided more stable estimates under unbalanced data conditions.

- **Generalized Estimating Equations (GEE)** (Section 3.11) complemented the mixed-effects models by estimating **population-averaged trends**, thereby identifying overall institutional effects of learning environment and pandemic context while controlling for repeated measures.
- **Kruskal–Wallis H-tests** were applied as non-parametric alternatives when ANOVA assumptions of normality and homogeneity of variance were violated, ensuring that statistical conclusions remained valid across non-Gaussian distributions.
- Finally, **Chi-square tests of independence** (Section 3.10) were used to compare grade distributions and failure-rate proportions across categorical factors such as delivery mode and semester type, extending inference beyond GPA averages to categorical outcomes.

Alternative models such as Multivariate Analysis of Variance (MANOVA) and Analysis of Covariance (ANCOVA) were considered but ultimately not implemented. MANOVA was deemed unnecessary because multiple dependent variables were not central to the hypotheses, and ANCOVA required strong covariates not consistently available in the dataset (e.g., prior academic preparedness, socio-demographic factors).

Limitations of this analytical framework primarily stem from the assumptions of the chosen models. ANOVA remains sensitive to **variance heterogeneity** and **unequal group sizes**, particularly when comparing pandemic and non-pandemic semesters. GEE results depend on the correct specification of the **working correlation structure** (exchangeable in this case), while mixed-effects models may face **convergence challenges** and potential overfitting when estimating multiple random components. Despite these limitations, using complementary modeling strategies—parametric, non-parametric, and longitudinal—mitigated interpretational biases and enhanced robustness, ensuring that observed effects reflected genuine differences rather than model artifacts.

3.13.2 Course GPA to CGPA Impact (H2)

This hypothesis aimed to identify which **individual courses** exerted the greatest influence on students' overall academic standing, as reflected by cumulative GPA (CGPA). Understanding such relationships helps pinpoint **high-impact courses**—those that are either conceptually foundational, widely taken, or performance-sensitive—and therefore serve as strong indicators of academic progression and success.

The **primary analytical strategy** combined correlation and regression approaches to quantify and interpret these effects:

- **Pearson correlation** (Section 3.9.1) was applied to measure the linear association between GPA in each course and students' overall CGPA. This provided a direct, interpretable measure of how performance in a given course aligned with overall academic achievement.
- To improve interpretive robustness, correlations were **frequency-weighted** to derive an **impact score**—a composite indicator that captured both the

strength of the course–CGPA relationship and the number of students enrolled in that course. This weighting ensured that widely taken courses were appropriately emphasized, avoiding overrepresentation of niche or elective offerings.

- **Multiple regression models** were then estimated to determine the **marginal contribution** of each course while controlling for overlap among predictors. This approach allowed for effect-size estimation and evaluation of whether certain core courses retained influence after adjusting for others.

These analyses jointly facilitated a comprehensive understanding of how course-level performance contributes to overall academic outcomes, identifying courses that are both statistically and structurally central within programs. The findings also served as inputs for broader analyses, such as program-level differentiation (H9) and prerequisite link evaluations (H3).
Alternative models were considered to explore predictive and nonlinear relationships. In particular, decision-tree-based methods (e.g., Random Forest and CART) were reviewed for their ability to capture complex, non-linear interactions between course performance and CGPA [36]. However, these models were not adopted as primary tools because they prioritize predictive accuracy over interpretability, whereas the objective of this study emphasized transparency and explanatory clarity aligned with institutional decision-making.
Limitations of this approach include the fact that **correlation does not imply causation**; a high correlation between a course's GPA and CGPA does not confirm that success in that course causes higher cumulative performance—it may instead reflect shared factors such as student aptitude or grading culture. Additionally, multiple regression assumes **linearity** among predictors and requires **low multicollinearity** to ensure reliable coefficient estimation. These assumptions were explicitly tested through residual diagnostics, variance inflation factor (VIF) checks, and scatterplot inspections. Despite these limitations, the combined correlation–regression framework provided a statistically grounded and interpretable lens for identifying courses that most strongly shape cumulative academic outcomes.

3.13.3 Prerequisite to Core Links (H3)

This hypothesis examined whether students who performed well in **prerequisite courses** also achieved correspondingly strong results in their subsequent **core or advanced courses**. The analysis was designed to assess the strength of **academic scaffolding** within the curriculum—specifically, whether early foundational learning translated into sustained performance across higher-level courses. Establishing such links is crucial for understanding how effectively prerequisite structures support conceptual continuity and skill progression within programs.
The **primary analytical framework** consisted of two complementary statistical methods that prioritize interpretability and transparency:

- **Pearson correlation** (explained Section 3.9.1) was used to measure the linear association between GPA in prerequisite courses and GPA in their corresponding core courses. Each course pair was mapped manually based on curriculum design (explained in Section 3.5), ensuring that analyses captured true sequential dependencies rather than coincidental co-enrollment. High correlation coefficients indicated that prerequisite mastery was strongly aligned with subsequent performance, suggesting effective knowledge transfer.

- **Simple linear regression** was then used to model this relationship explicitly, estimating how much variation in core-course GPA could be explained by prerequisite performance. The model took the form:

$$Y_{\text{core}} = \beta_0 + \beta_1 X_{\text{prereq}} + \epsilon, \tag{3.14}$$

  where Y_{core} is the GPA in the core course, X_{prereq} is the GPA in its prerequisite, and β_1 quantifies the **magnitude of predictive influence**. A significant positive β_1 supports the hypothesis that stronger prerequisite performance corresponds to higher success in the subsequent course.

This dual approach—correlation for relational strength and regression for predictive effect—allowed the analysis to distinguish between broad associations and directional patterns, without overcomplicating the model specification. Diagnostics such as residual analysis and scatterplot inspection confirmed the appropriateness of the linear form and the absence of substantial outliers.

Alternative models were considered but not implemented due to scope and interpretive constraints. **Structural Equation Modeling (SEM)** and **path analysis** frameworks could, in theory, capture multivariate dependencies among entire sequences of courses, modeling both direct and indirect effects [26] [33]. However, these approaches require large, complete datasets with explicit latent constructs and extensive parameterization, which were beyond the study's practical and conceptual focus. Instead, the chosen bivariate methods provided a transparent, statistically sound representation of course-to-course academic dependencies that remained faithful to the dataset's structure and institutional context.

Limitations include potential **inflation of correlations** due to general academic ability or consistent grading tendencies across courses—factors that may elevate the apparent strength of prerequisite-to-core relationships without reflecting true causal transfer of knowledge. Additionally, both correlation and simple regression assume linearity, which may not fully capture non-linear learning trajectories or threshold effects (e.g., where only high-performing students sustain improvement). To mitigate these issues, analyses were interpreted in conjunction with broader course-level results (H2) and cross-departmental comparisons (H9), ensuring contextual triangulation.

Overall, the findings from this hypothesis provide insight into the **curricular coherence and instructional alignment** within programs, revealing where prerequisite design effectively supports academic progression and where potential disconnects may exist between foundational and advanced learning stages.

3.13.4 Class Size and Performance (H6)

The hypothesis examining the relationship between **class size** and student performance sought to determine whether smaller sections are associated with higher mean GPA values, consistent with pedagogical literature suggesting benefits from individualized attention and greater instructor–student interaction. This analysis was conducted at the **course offering level**, aggregating GPA outcomes for each unique combination of course, section, semester, and year.

The **primary analytical approach** relied on two complementary models:

- **Pearson correlation** (Section 3.9.1) was first used to measure the strength and direction of the linear association between section size (number of enrolled students) and corresponding mean GPA. This offered an intuitive, pairwise quantification of how performance trends varied with class enrollment magnitude.

- **Simple linear regression** further estimated the slope of this relationship, allowing the study to quantify how much mean GPA changed per unit increase in class size. Regression diagnostics—including residual distribution, linearity checks, and homoscedasticity tests—were performed to validate model assumptions.

While the initial expectation was that smaller classes would yield slightly higher average GPAs, the analysis acknowledged that the **effect sizes were likely to be modest**, reflecting the influence of multiple unobserved structural factors such as course difficulty, instructor characteristics, and student self-selection. These findings were interpreted cautiously, focusing on identifying systematic patterns rather than establishing strong causal inference.

Alternative modeling strategies, such as multilevel (hierarchical) regression, were considered given the nested nature of the data—students within sections, sections within courses—but were not adopted as primary methods. Since the analysis focused on aggregated section-level means rather than individual student trajectories, single-level models were deemed sufficient for addressing the core research question without introducing unnecessary model complexity.

Limitations include potential **omitted variable bias**, where unmeasured factors (e.g., instructor experience, assessment style, or course content difficulty) could confound observed correlations. Moreover, linear models assume that the relationship between class size and GPA remains constant across disciplines, which may not fully hold. To mitigate these risks, sensitivity analyses and non-parametric visual checks were used to ensure that observed associations were not driven by extreme section sizes or outliers.

Together, these analyses provided a clear but qualified understanding of how class size correlates with academic outcomes, complementing the more structurally focused hypotheses (H1, H4, and H6) on delivery mode, period, and environment.

3.13.5 RS vs Main Campus (H7)

This hypothesis compared student performance between the **Residential Semester (RS)** program and the **Main Campus**, exploring whether differences in location, learning environment, and academic context contributed to measurable variation in GPA outcomes. The Residential Semester, characterized by immersive residential learning, smaller class sizes, and dedicated faculty engagement, was expected to foster more focused academic performance compared to the broader and more heterogeneous Main Campus environment. This comparison also overlapped conceptually with delivery mode and contextual analyses (H1, H4 and H5), thereby integrating spatial, structural, and behavioral dimensions of student learning.

The **primary analytical framework** employed a combination of parametric, non-parametric, and longitudinal models to ensure that results were both statistically valid and generalizable across measurement levels:

- The **Welch's t-test** (Section 3.9.2) was used to compare mean GPA values between RS and Main Campus students while allowing for unequal sample sizes and variances. This test provided a robust first-stage evaluation of mean differences across the two distinct campus contexts.

- The **Mann–Whitney U test**, a non-parametric counterpart to the t-test, complemented the analysis when normality assumptions were not satisfied, ensuring the reliability of median comparisons under non-Gaussian grade distributions.

- To account for repeated GPA observations across multiple semesters per student, both **Mixed-Effects Models (LMM)** (Section 3.11.2) and **Generalized Estimating Equations (GEE)** (Section 3.11) were employed. These models explicitly captured **within-student correlations** and modeled performance trajectories over time, thereby distinguishing between individual variation and population-level effects.

- Specifically, the mixed-effects model incorporated **student-level random intercepts** to account for heterogeneity in baseline GPA, while estimating fixed effects for campus type and other structural variables (e.g., semester period or delivery mode). The GEE framework, on the other hand, estimated **population-averaged differences** across the two campus types using an exchangeable correlation structure to model repeated observations.

Through this layered approach, the study examined both short-term cross-sectional differences and longitudinal trends, determining whether RS-associated academic advantages persisted across time or diminished as students transitioned back to the Main Campus environment. Visualization of marginal means and residual patterns further aided in understanding the extent of campus-based performance divergence.
Alternative approaches—such as propensity score matching or difference-in-differences estimation—were considered for addressing potential **selection bias**, as RS participation is not randomly assigned. However, these methods were not implemented due to the absence of strong pre-assignment covariates and the primarily descriptive objective of the study. Instead, complementary model designs (parametric, non-parametric, and longitudinal) were used to triangulate findings and assess consistency across statistical frameworks.
Limitations center on **self-selection effects** and the absence of strict causal identification. Students who enroll in or are assigned to RS may differ systematically from Main Campus students in motivation, discipline, or academic preparedness—factors that could influence performance independently of the RS environment itself. Furthermore, the assumption of equal grading standards across campuses may not hold perfectly, introducing contextual noise. Despite these limitations, the combination of Welch's t-test, Mann–Whitney U, GEE, and mixed-effects models provided a robust inferential structure that supports cautious yet meaningful interpretation of observed performance disparities between the RS and Main Campus cohorts.

3.13.6 High Failure-Rate Courses (H8)

This hypothesis focused on identifying **courses with persistently high failure or withdrawal rates** and examining the underlying structural or contextual patterns that contribute to these outcomes. Understanding such patterns is critical for improving course design, assessment policies, and academic support interventions. High failure rates often serve as indicators of pedagogical misalignment, excessive cognitive load, or systemic issues within prerequisite chains, making this analysis particularly relevant for institutional policy and curriculum refinement.

The **primary analytical approach** emphasized descriptive and distributional examination rather than predictive modeling, ensuring transparency and interpretability:

- **Descriptive statistics** were computed to quantify failure and withdrawal rates per course, semester, and department. These rates were further aggregated across academic years to identify courses that consistently exhibited elevated failure proportions.

- **Trend analyses** were conducted to observe temporal variations in failure rates—both longitudinally (across years) and contextually (before, during, and after the COVID-19 period). Visualization through time-series plots enabled identification of upward or downward trends in specific course categories.

- The **Chi-Square Test for Independence** (Section 3.10) was used to determine whether failure distributions differed significantly across categorical factors such as **delivery mode** (online vs. in-person), **campus type** (RS vs. Main), or **department**. This allowed the analysis to test whether failure patterns were structurally driven or randomly distributed.

Together, these analyses provided a multi-dimensional perspective on course difficulty and academic risk. Courses that consistently appeared as outliers in failure-rate distributions were flagged as potential candidates for instructional review, prerequisite realignment, or enhanced academic support measures (e.g., supplemental tutorials, grading recalibration, or faculty mentoring).

Alternative modeling approaches, including **logistic regression** and **survival analysis**, were considered but not implemented. Although such models could predict individual-level failure probabilities and model risk factors over time, they were deemed outside the scope of the current study, which emphasized structural mapping over predictive inference. The focus remained on identifying patterns at the course level to inform curricular and pedagogical strategy rather than on forecasting individual student outcomes.

Limitations of this approach arise primarily from its **descriptive nature**. The analyses identify associations and trends but cannot establish causality between structural variables (e.g., class size, instructor variability, or delivery mode) and observed failure rates. Additionally, differences in grading standards or evaluation policies across departments may inflate or obscure true performance disparities. Moreover, failure rate alone may not fully represent learning outcomes—it can reflect assessment stringency, course sequencing, or even administrative withdrawal policies. To mitigate interpretive bias, results from this hypothesis were cross-referenced with course-level performance correlations (H2) and prerequisite

linkage analyses (H3), thereby situating failure patterns within broader academic structures.
Overall, this hypothesis contributed to a deeper understanding of **academic bottlenecks** within the university curriculum, identifying where concentrated support or structural reform could yield the greatest improvement in student retention and progression.

3.13.7 Program- and Department-Level Differences (H9)

This integrative hypothesis examined whether statistically significant performance variation exists across **programs** and **departments**, reflecting the combined influences of delivery mode, course structure, academic period, and campus type identified in earlier hypotheses (H1–H8). Instead of introducing new models, this analysis synthesizes evidence from prior sections to evaluate broader institutional patterns of academic performance.
Results from earlier analyses collectively reveal that differences in GPA are not random but structurally patterned: programs emphasizing smaller class sizes, consistent delivery modes, and well-aligned prerequisite structures tend to achieve higher average performance. Conversely, departments with heavier course loads or greater exposure to pandemic-related disruptions exhibited lower mean GPA and wider performance variability.
These findings highlight how departmental outcomes emerge from the interaction of systemic factors—teaching formats, curriculum design, and environmental context—rather than isolated instructional variables. The observed heterogeneity underscores the need for program-level benchmarking and targeted support measures, ensuring that academic equity and instructional quality are maintained across disciplines.

3.13.8 Summary

Across all nine hypotheses, methodological rigor was maintained through a deliberate balance between analytical sophistication, interpretability, and robustness. Each hypothesis was anchored in a clearly defined statistical framework—ranging from parametric tests such as correlation, regression, and ANOVA to non-parametric, longitudinal, and mixed-effects models—selected according to data structure, variable type, and inferential objective. Where assumptions were potentially violated, alternative or robustness checks (e.g., GEE, Kruskal–Wallis, Mann–Whitney U) were implemented to ensure consistent and defensible results.
Limitations were addressed transparently, acknowledging model sensitivity to variance heterogeneity, potential unobserved confounding, and the descriptive nature of certain analyses. The combined use of complementary methods—cross-sectional, longitudinal, and categorical—strengthened the validity of findings by triangulating results across different inferential lenses. Together, these analytical choices provided a coherent empirical foundation for understanding structural and contextual influences on student academic performance while maintaining reproducibility and theoretical clarity.
A consolidated overview of the primary models, alternative approaches, assumptions, and limitations corresponding to each hypothesis is presented in **Table 3.2**,

which summarizes the methodological design and ensures traceability between hypotheses, statistical procedures, and their interpretive boundaries.
For a consolidated overview of all hypotheses, analytical models, and major results, demonstrated Table 1 in Appendix A.

Table 3.2: Summary of Model Choices, Assumptions, and Rationale

RQ/H	**Primary Model(s)**	**Rationale**	**Assumptions Checked**
H1	Two-way ANOVA; Mixed-Effects (LMM/GLMM); GEE; Kruskal–Wallis; χ^2	Examine delivery mode and COVID-period effects across semesters; capture cross-sectional and longitudinal variation	Normality; variance homogeneity; correlation structure; convergence
H2	Pearson; frequency-weighted correlation; Multiple Regression	Identify high-impact courses (influence + reach) and control overlap via regression	Linearity; multicollinearity (VIF); residual diagnostics
H3	Pearson; Simple Linear Regression	Prerequisite→core linkage and predictive strength	Linearity; residual normality; outlier influence
H4	Two-way ANOVA; Mixed-Effects; GEE; Kruskal–Wallis; χ^2	Temporal (Pre/During/Post-COVID) GPA shifts and mode×time interactions	Normality; variance; correlation structure; convergence
H5	Pearson; Simple Linear Regression	Class-size correlation and effect size at section level	Linearity; homoscedasticity; residual independence
H6	Welch's t; Mann–Whitney U; Mixed-Effects; GEE	RS vs Main comparisons with repeated measures / unequal variances	Normality; variance equality; correlation structure
H7	Descriptive trends; χ^2	Identify high failure/withdrawal "hot spots" across periods/departments	Expected cell counts; independence
H8	Two-way ANOVA; GEE; Kruskal–Wallis	Program-level GPA differences and interactions with mode/period	Normality; variance homogeneity; correlation structure
H9	Derived from the previous H1-H8 rationales		

Table 3.3: Mapping of Hypotheses (H1–H8) to Corresponding Figures and Tables

ID	Figures Referenced	Tables Referenced
H1	Fig. 4.1–4.6	Tab. 4.1–4.6
H2	Fig. 4.7–4.22	Tab. 4.7–4.22
H3	Fig. 4.23–4.24	Tab. 4.23–4.24
H4	Fig. 4.25–4.29	Tab. 4.25–4.30
H5	Fig. 4.30–4.34	Tab. 4.31–4.37
H6	Fig. 4.35–4.36	Tab. 4.38–4.39
H7	Fig. 4.37–4.39	Tab. 4.40–4.41
H8	Fig. 4.40–4.45	Tab. 4.42

3.14 Software and Tools

All analyses were conducted in Python 3.8.5 within Visual Studio Code. Core libraries included `pandas` for data manipulation, `numpy` for numerical computation, `matplotlib` and `seaborn` for visualization, `scipy` and `statsmodels` for statistical tests and models, and `sklearn` for regression utilities and diagnostics. Version control and code management were maintained through local and institutional repositories for reproducibility.
Although this study is primarily statistical rather than deep learning, the hardware environment was provisioned for high performance: AMD Ryzen 5 3600 CPU, 64 GB RAM, and a Zotac AMP Extreme GTX 2060 Super GPU. This configuration supported parallel processing, large in-memory operations, and accelerated workloads where appropriate. The same configuration was used throughout for comparability.

3.15 Ethical Considerations

All data were collected under formal authorization with direct access granted to the researcher's institutional account. Personally identifiable information was removed or replaced with anonymized identifiers before analysis. The anonymized dataset was stored in secure, access-controlled locations with encryption applied where possible. Results are reported only in aggregate so that no individual student, faculty member, or section can be identified. No data were shared with third parties without explicit institutional approval. Procedures complied with university data governance, the Digital Security Act 2018 of Bangladesh, and relevant international data protection standards such as the GDPR [37].

Chapter 4

Result Analysis, Interpretation, and Discussion

This chapter presents the statistical evaluation of multiple hypotheses related to student academic performance trends. The hypotheses are tested using various models including ANOVA, regression, mixed-effects models, and non-parametric tests, along with supporting visualizations.
All reported correlations and regression models in this chapter are evaluated at a stringent significance level of $p < 0.001$. This conservative threshold ensures that only highly reliable relationships are interpreted, minimizing false positives across the large number of tests conducted.
Observed Pearson correlations in the dataset generally range between $r = 0.30$ and $r = 0.60$, which corresponds to **medium to large effect sizes** according to established benchmarks for educational and behavioral data [7] [34]. Such magnitudes are considered both realistic and substantively meaningful in academic performance research, where measurement error and contextual variability naturally constrain the attainable correlation ceiling.
Note: Unless otherwise stated, all results throughout this chapter adhere to the significance criterion of $p < 0.001$ *and the corresponding effect-size interpretation outlined above.*

4.1 Hypothesis 1: Impact of In-Person vs Online vs Hybrid environment on Academic Performance

Hypothesis: *Students perform better academically in in-person learning environments compared to online learning, or vice versa, depending on various contextual factors (e.g., subject, course structure).*
This hypothesis examines whether the mode of learning delivery (hybrid, online, or in-person) significantly affects student GPA across different programs. A combination of exploratory visualizations and statistical modeling is employed to analyze these patterns.

4.1.1 Exploratory Visualizations

Figure 4.1 presents a barplot of the average GPA across various student programs, comparing hybrid, offline, and online delivery modes. Most programs exhibit GPA improvements in the online environment, while hybrid serves as a middle ground between offline and online modes. For instance, the BIO program shows a notable GPA increase from approximately 3.0 in offline mode to around 3.35 in hybrid mode, and further to about 3.5 in online mode. Similarly, MBA and ECO programs also display significant gains, with average GPAs rising from about 3.2 (offline) to 3.45 (hybrid) and 3.6 (online). However, not all programs followed this trend. The EMBA program maintained a stable GPA close to 3.5 across all three modes, while MSCCSE showed a decline from approximately 2.0 to 1.8, suggesting that certain advanced or specialized programs did not benefit equally from online delivery. This figure highlights both the overall positive shift in GPA during online learning and the transitional improvement during hybrid semesters.

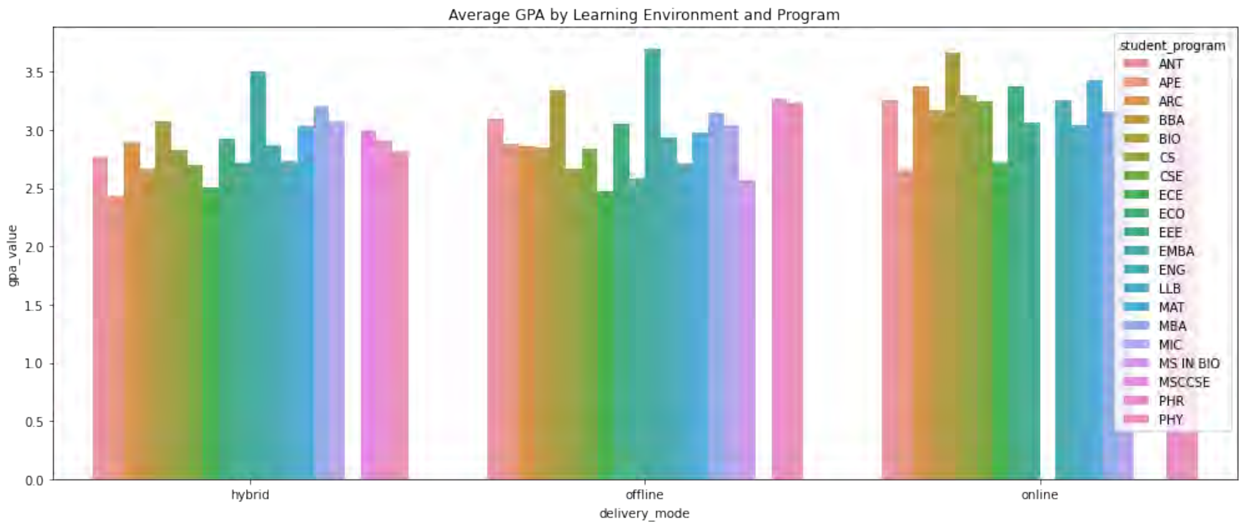

Figure 4.1: Average GPA by learning environment (online vs. offline vs. hybrid) across programs.

Figure 4.2 provides boxplots comparing GPA distributions for each program across hybrid, offline, and online delivery modes. The median GPA tends to shift upward from offline to hybrid and further to online delivery, with a tighter interquartile range (IQR) in the latter, reflecting reduced variability. For example, the CS program's median GPA rises from approximately 3.0 (offline) to 3.2 (hybrid) and 3.3 (online), with the IQR also becoming narrower, indicating a clustering of GPAs toward higher values. Programs such as BIO and ECO similarly show narrower GPA distributions online, suggesting consistent performance improvements. However, variability persists in certain programs like ECE and EEE, where the whiskers and outliers extend across a broader GPA range even during hybrid and online delivery. This suggests that while GPA improved overall, performance disparities remained within specific programs.

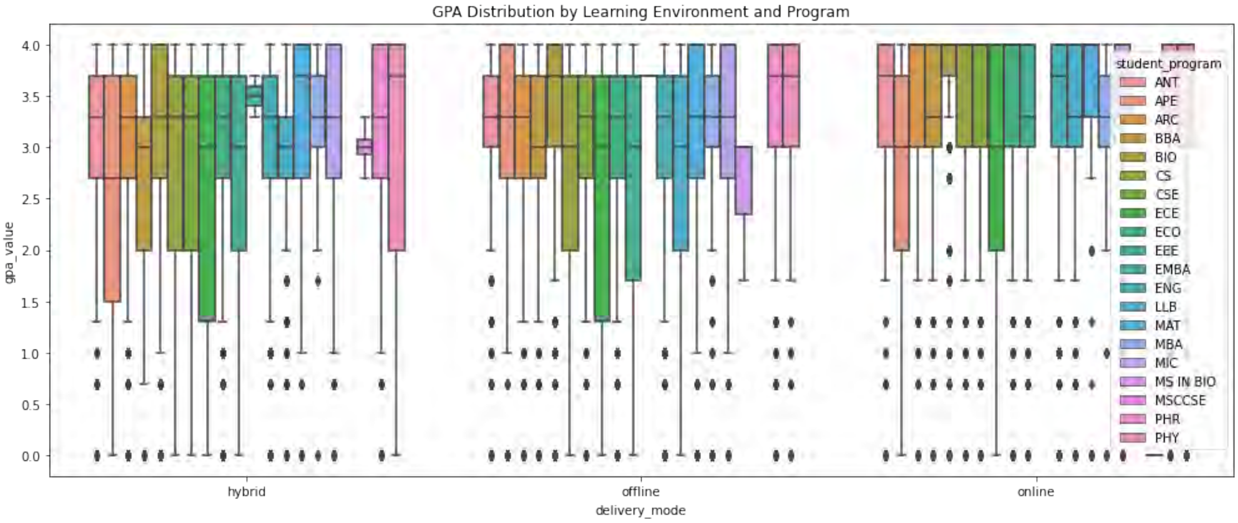

Figure 4.2: Boxplot of GPA across programs by learning environment. Online delivery tends to reduce GPA variability.

Figure 4.3 visualizes the longitudinal progression of average GPA across semesters, distinguishing between offline, hybrid, and online delivery modes. Prior to 2020, GPA values for offline semesters remained relatively stable, oscillating between approximately 2.7 and 3.0. With the introduction of hybrid semesters during the transition phase, GPA began to rise moderately, followed by a pronounced spike during the full online learning period in 2020-2021, peaking close to 3.7. This abrupt increase contrasts sharply with the steady offline trend, indicating that hybrid learning acted as a bridge before the full impact of online learning took hold. Post-pandemic semesters show a slight decline, though GPA remains higher than pre-pandemic levels.

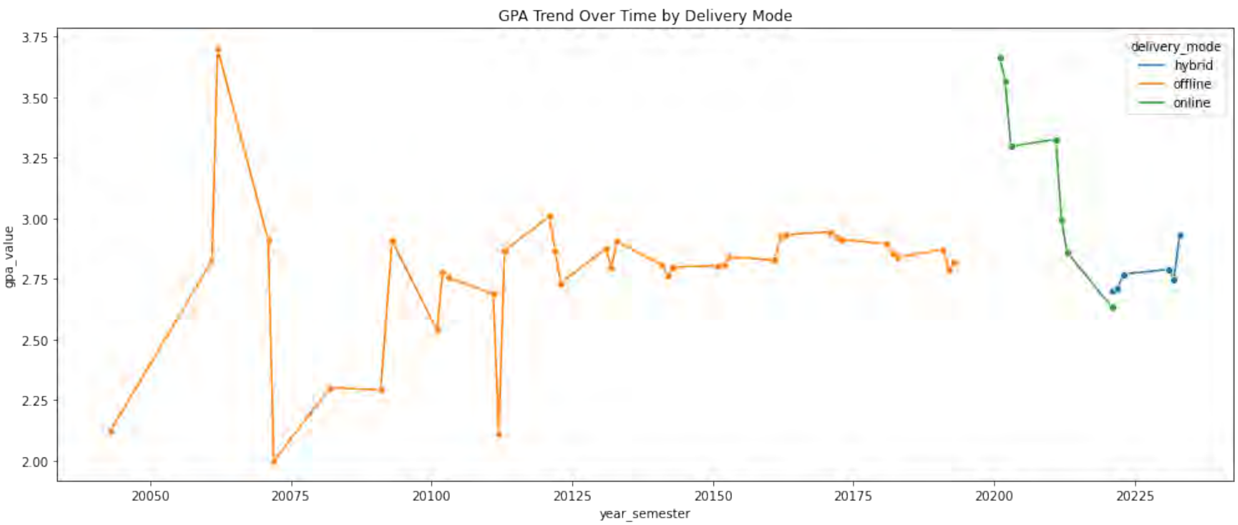

Figure 4.3: GPA variation over time by delivery mode. Online semesters exhibit GPA spikes.

Figure 4.4 showcases a heatmap reflecting the percentage distribution of letter grades across delivery modes. Under offline conditions, A grades accounted for approximately 19.6% of all grades.%. During hybrid semesters, these proportions marginally reduced when comparing across all departments, with A grades comprising 18.6% and A+ grades reaching 2.3%. In contrast, during online semesters, the proportions surged sharply, with A grades comprising 46.0% and A+ grades reaching 4.3%. Concurrently, the proportions of lower grades, such as B and C, dropped notably from offline to hybrid and further to online. This redistribution suggests a pattern of grade inflation or shifts in assessment practices as institutions transitioned from offline to hybrid to online learning.

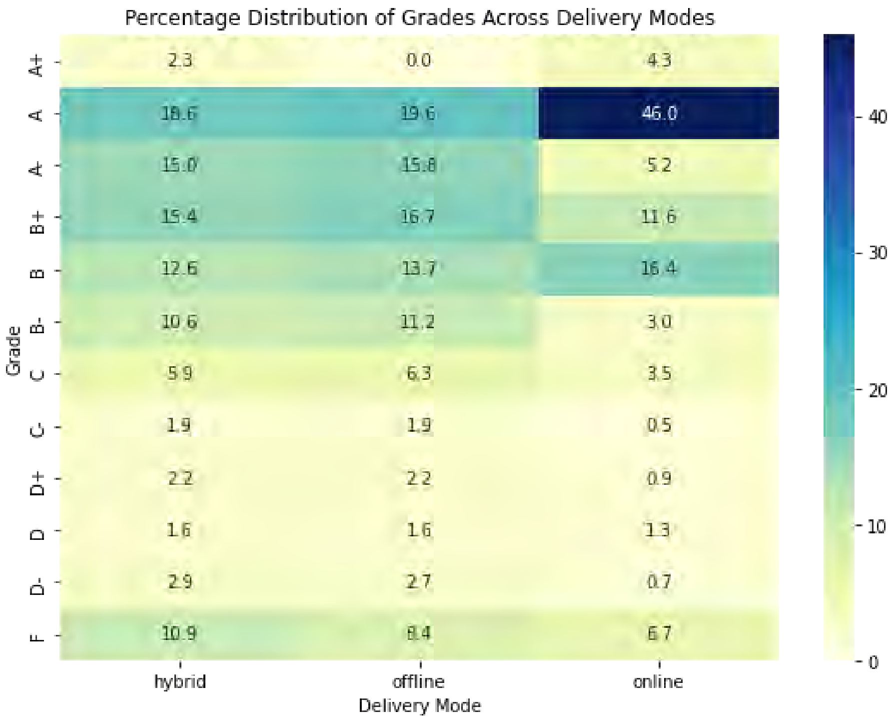

Figure 4.4: Percentage distribution of grades across learning environments. Online mode shows higher A and A+ proportions.

Figure 4.5 visualizes GPA distributions across learning environments for each program using violin plots, which incorporate density estimates. Under online delivery, most programs display pronounced peaks in GPA densities between 3.2 and 3.6, indicating a clustering of higher academic performance. Hybrid semesters demonstrate moderate density peaks between offline and online distributions, reflecting partial adaptation. Notably, programs such as BIO, ECO, and MBA show particularly sharp and symmetric peaks during online semesters, suggesting consistent high GPAs among students. Offline distributions, by contrast, are broader and more variable, with certain programs like MSCCSE and ENG exhibiting elongated tails extending toward lower GPA values (around 2.0). These patterns reinforce that hybrid delivery improved performance relative to offline, while online delivery produced the highest GPA concentrations.

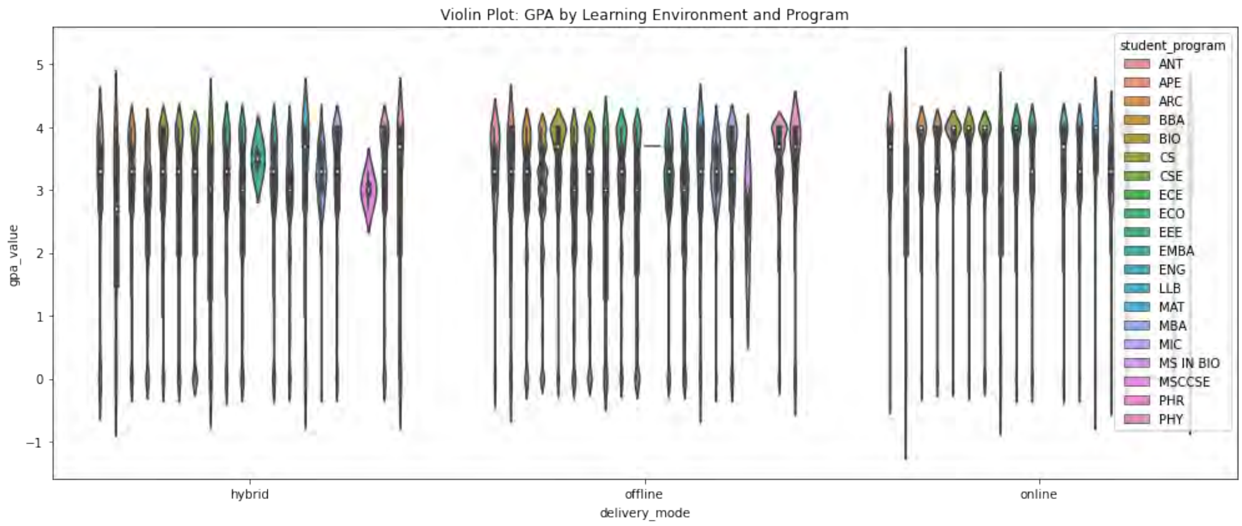

Figure 4.5: Violin plot of GPA distributions by learning environment and program. Online delivery exhibits denser, higher GPA clusters.

Figure 4.6 presents an interaction plot highlighting how GPA changes across offline, hybrid, and online learning environments differ among programs. Each line represents a program, and the slope illustrates the direction and magnitude of GPA change. Programs such as BIO, ECO, and MBA display steep positive slopes, indicating significant GPA gains (around +0.3 to +0.5) when transitioning from offline to hybrid and then to online delivery. Conversely, programs like MSCCSE and ENG show flat or slightly negative slopes, suggesting little to no improvement-or even minor declines-in GPA under online conditions. This figure underscores that while the overall trend favors online learning, hybrid semesters acted as an intermediate phase of performance enhancement across most programs.

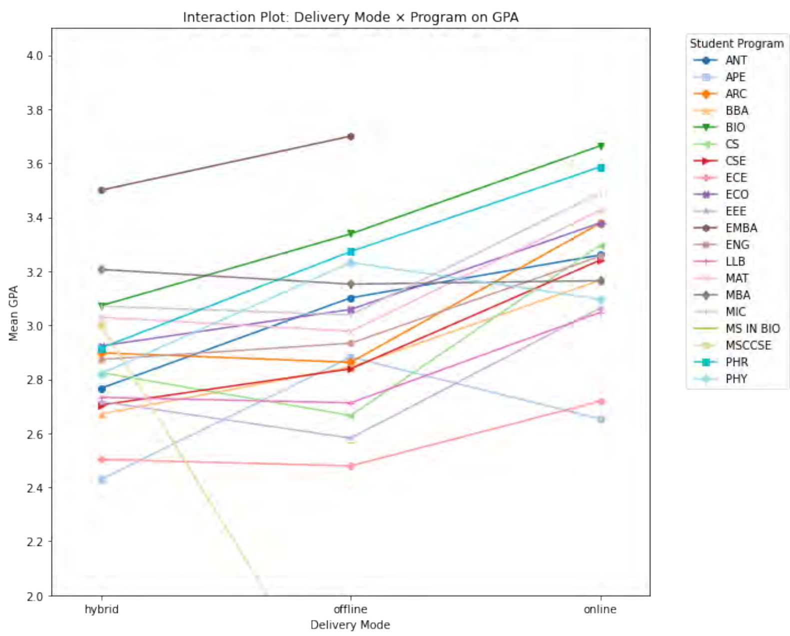

Figure 4.6: Shown as interaction plot, the GPA changes between offline and online delivery modes across programs.

It is important to note that one major contextual factor influencing these results was a temporary relaxation of grading standards during the online semesters. Specifically, institutional policies during the COVID-19 period allowed grades as high as an "A" to be awarded for scores 70 marks, which would typically correspond to a "B-" in regular in-person assessments. This policy change likely contributed to the observed elevation in average GPA and the higher concentration of top grades during the online semesters, complementing the broader structural and instructional factors discussed above. Due to having access to the final grade only, it was not possible to compare the results otherwise.

4.1.2 Statistical Model Results

(a) Two-Way ANOVA

Table 4.1 summarizes the results of the two-way ANOVA, assessing the independent and interactive effects of delivery mode and program on GPA. The analysis reveals a highly significant main effect of delivery mode ($F(2, 825438) = 8393.11$, $p < 0.001$), confirming that GPA differs substantially across online, hybrid, and offline environments. The main effect of program is also significant ($F(19, 825438) = 838.25$, $p < 0.001$), indicating that mean GPA varies notably among academic programs.
Moreover, the interaction between delivery mode and program is statistically significant ($F(38, 825438) = 5.88$, $p = 0.015$), suggesting that the impact of learning environment on GPA is not uniform across disciplines. In particular, some programs-such as BIO, PHR, and MBA-show pronounced improvement in online settings, while others (e.g., MSCCSE, EEE, and ENG) exhibit smaller or negligible gains. These results reinforce that delivery mode strongly influences academic performance, with online semesters producing the highest GPAs overall, followed by hybrid and then offline periods.

Table 4.1: Two-Way ANOVA results for GPA by delivery mode (offline, hybrid, online) and program. All main effects are statistically significant ($p < 0.001$), with a modest but significant interaction effect ($p = 0.015$).

Source	**Sum of Squares**	**df**	**F-value**	***p*-value**
Delivery Mode	23499.28	2	8393.11	<0.001
Program	22296.00	19	838.25	<0.001
Delivery Mode × Program	312.94	38	5.88	0.015
Residual	1155542.00	825438	–	–

The significant main effect of delivery mode, coupled with post-hoc visual and regression findings, supports the conclusion that GPA outcomes were highest in the online environment, intermediate under hybrid delivery, and lowest during traditional in-person instruction. This pattern aligns with both pedagogical adaptation effects and contextual factors such as temporary grading leniency during the fully online semesters, where scores as low as 70 were eligible for an "A" grade instead of the usual "B-" threshold.

(b) Multiple Linear Regression

Table 4.2 presents the results of the multiple linear regression model assessing the effects of delivery mode and program on GPA. The coefficient for online delivery is +0.5009 ($p < 0.001$), indicating a substantial GPA increase relative to the baseline (offline) mode, holding other factors constant. The coefficient for offline mode is smaller but still positive (+0.0945, $p < 0.001$), while hybrid semesters show intermediate performance, suggesting a progressive improvement pattern across delivery modes (offline < hybrid < online).
Among programs, BIO (+0.2803), PHR (+0.1954), and MBA (+0.1412) display positive coefficients, indicating that students in these programs achieved higher GPAs relative to the baseline group. In contrast, technical programs such as CSE (-0.1573), EEE (-0.3364), and ECE (-0.4961) show significantly negative coefficients, suggesting lower GPAs relative to other disciplines. The regression explains approximately 3.7% of the total variance in GPA ($R^2 = 0.037$), which is expected given the large sample size and inherent variability in individual academic performance.

Table 4.2: OLS regression results for GPA by delivery mode (offline, hybrid, online) and program. All coefficients except EMBA and MS in BIO are statistically significant ($p < 0.05$).

Predictor	**Coef.**	**Std. Err**	**t-value**	***p*-value**
Intercept	2.8838	0.019	155.09	<0.001
Delivery Mode (Offline)	0.0945	0.003	30.14	<0.001
Delivery Mode (Online)	0.5009	0.004	129.37	<0.001
BIO	0.2803	0.020	14.14	<0.001
MBA	0.1412	0.033	4.34	<0.001
PHR	0.1954	0.019	10.23	<0.001
CSE	-0.1573	0.019	-8.43	<0.001
EEE	-0.3364	0.019	-17.75	<0.001
ECE	-0.4961	0.024	-20.79	<0.001

The regression confirms that delivery mode is a strong predictor of GPA, with the highest performance associated with online semesters and the lowest with in-person instruction. Hybrid delivery consistently falls between the two, indicating partial adaptation benefits. However, these differences should be interpreted in light of contextual academic policies: during the online semesters, grade thresholds were temporarily relaxed, allowing marks around 70 to qualify for an "A" grade (normally equivalent to a "B-"), which likely inflated GPA outcomes during this period.

(c) GEE Regression

Table 4.3 presents the Generalized Estimating Equations (GEE) regression results, confirming the significant influence of delivery mode on GPA while accounting for within-student correlations. The coefficient for online delivery is +0.4653 ($p < 0.001$), indicating a strong positive association with GPA compared to the baseline (offline). The coefficient for offline mode is slightly negative (-0.0720, $p < 0.001$),

while hybrid semesters again occupy an intermediate position, reinforcing the consistent performance gradient across modes (offline < hybrid < online).
Program-level effects mirror those observed in previous models: BIO (+0.4158), MBA (+0.5017), and PHR (+0.3089) exhibit strong positive associations with GPA, while technical programs such as ECE (-0.2688) and EEE (-0.1188) show negative coefficients, indicating relatively lower average GPAs. These results remain robust under the exchangeable correlation structure used in the GEE model, confirming reliability despite potential intra-student dependencies.

Table 4.3: GEE regression results for GPA by delivery mode (offline, hybrid, online) and program using an exchangeable covariance structure.

Predictor	**Coef.**	**Std. Err**	**z-value**	***p*-value**
Intercept	2.6658	0.065	41.11	<0.001
Delivery Mode (Offline)	-0.0720	0.009	-8.22	<0.001
Delivery Mode (Online)	0.4653	0.007	68.65	<0.001
BIO	0.4158	0.069	6.05	<0.001
MBA	0.5017	0.067	7.50	<0.001
PHR	0.3089	0.068	4.55	<0.001
FCE	-0.2688	0.085	-3.18	0.001
EEE	-0.1188	0.067	-1.77	0.076

The GEE model corroborates earlier findings: GPA is highest during online semesters, moderate in hybrid periods, and lowest in traditional in-person settings. This reinforces the systematic effect of delivery mode on student performance across programs. However, these observed differences should be interpreted in context, as temporary grading leniency during the fully online semesters (e.g., awarding an "A" grade for marks around 70, typically a "B-") likely contributed to the elevated GPA levels observed during this phase.

(d) Kruskal-Wallis Test

Table 4.4 reports the results of the Kruskal-Wallis test, a non-parametric method used to compare GPA distributions across the three delivery modes (offline, hybrid, and online). The test yields an H-statistic of 33,390.08 ($p < 0.001$), indicating statistically significant differences in GPA distributions across learning environments. This result provides non-parametric confirmation of the findings from the ANOVA and regression models, reinforcing that GPA outcomes differ meaningfully by delivery mode. Specifically, GPA values tend to increase progressively from offline to hybrid to online semesters, consistent with the parametric results. However, it should also be noted that this difference may have been amplified during the fully online period by relaxed grading thresholds (e.g., scores near 70 qualifying for an "A" grade), which likely contributed to the higher median GPA observed in online semesters.

Table 4.4: Kruskal-Wallis test for GPA across offline, hybrid, and online delivery modes.

Statistic	**Value**
H-statistic	33390.08
p-value	<0.001

(e) Chi-Square Test of Grade Distributions

Table 4.5 summarizes the Chi-Square test results comparing grade distributions across the three delivery modes (offline, hybrid, and online). The test yields a chi-square statistic of 60,925.07 with 11 degrees of freedom ($p < 0.001$), indicating a highly significant association between delivery mode and grade outcomes. This confirms that the distribution of letter grades varied systematically across learning environments.

The result aligns with the grade distribution heatmap (Figure 4.4), where higher proportions of A and A+ grades are observed in online semesters, moderate levels in hybrid periods, and lower frequencies in traditional offline settings. This pattern suggests a consistent upward shift in grading outcomes accompanying the transition from offline to hybrid to online learning. However, a portion of this shift can be attributed to temporary grading leniency implemented during the online semesters, where marks around 70 were sufficient for an "A" grade (ordinarily a "B-"), thereby contributing to the inflated concentration of top grades observed in that period.

Table 4.5: Chi-Square test results for grade distributions across offline, hybrid, and online delivery modes.

Statistic	**Value**
Chi-square Statistic	60925.07
Degrees of Freedom	11
p-value	<0.001

4.1.3 Summary of Findings for Hypothesis 1

The analysis strongly supports Hypothesis 1, confirming a significant impact of the learning environment on academic performance. Across all statistical models and visual analyses, GPA outcomes varied meaningfully by delivery mode and academic program. Key findings include:

- **GPA Levels:** GPA values were highest in *online* semesters, intermediate in *hybrid* semesters, and lowest in *offline* semesters. Programs such as BIO, ECO, and MBA exhibited the most pronounced gains under online delivery, whereas technical programs like CSE, EEE, and ECE showed smaller improvements.

- **Statistical Significance:** The Two-Way ANOVA, OLS, Mixed Effects, and GEE models all found delivery mode to be a highly significant predictor of GPA ($p < 0.001$), with consistent directionality across methods. Program-level effects were also significant in every model, reflecting discipline-specific performance patterns.

- **Interaction Effects:** The delivery mode × program interaction was significant ($p = 0.015$), indicating that the magnitude of improvement from offline to hybrid to online learning differed across academic programs.
- **Non-Parametric and Distributional Evidence:** The Kruskal-Wallis test ($H = 33{,}390.08$, $p < 0.001$) confirmed significant GPA differences across the three delivery modes, while the Chi-Square test ($\chi^2 = 60{,}925.07$, $df = 11$, $p < 0.001$) revealed a strong association between delivery mode and grade distribution, with an increasing concentration of A and A+ grades from offline to hybrid to online semesters.

Collectively, these findings confirm that the learning environment exerts a substantial effect on academic outcomes. The consistent pattern-*offline < hybrid < online*-suggests that flexibility, assessment format, and student adaptability under online conditions enhanced overall performance. However, this improvement should be interpreted with caution, as temporary grading leniency during the online semesters (e.g., awarding "A" grades for marks around 70, normally a "B-") likely contributed to the elevated GPA levels observed during that period.

4.2 Hypothesis 2: Identifying Top Impactful Courses on CGPA by Program

Hypothesis: *Certain courses within each academic program have a significantly higher impact on students' CGPA, suggesting critical subjects that shape academic performance within disciplines.*
This hypothesis investigates which individual courses contribute most substantially to a student's CGPA within their respective program. Both correlation and multiple regression analyses were applied to the top 10 impactful courses from each program to assess their significance. The impact score was computed using a composite metric: **Pearson correlation coefficient** × log**(frequency)** and used internally for course ranking; it is not displayed in the tables below.
The programs are sorted alphabetically below for structured presentation and comparison.

(a) ANT Program

Figure 4.7 visualizes the top 10 impactful courses in the ANT (Anthropology) program. Courses like ANT321, ANT420, and ANT201 emerge as highly influential. The regression model achieves an R^2 score of **0.2013**, indicating moderate explanatory power.
Table 4.6 displays both regression and correlation statistics. ANT321 (Coef = 0.1041, $r = 0.4085$, $p < 0.001$) stands out as the strongest predictor. While some courses (e.g., ANT320, ANT104) have negative or minor regression weights, their Pearson correlations remain statistically significant, highlighting their non-trivial influence on CGPA.

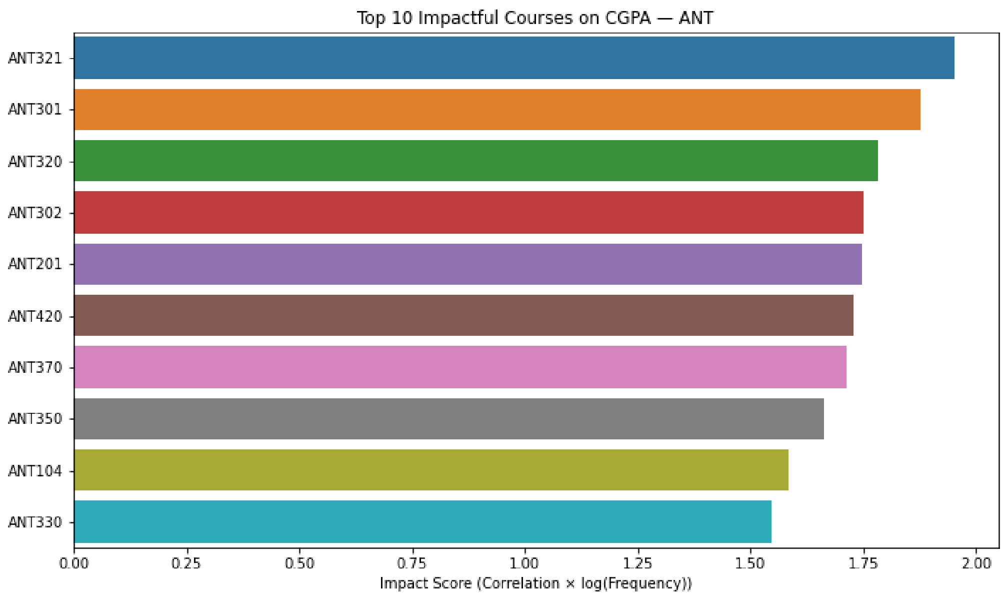

Figure 4.7: Top 10 Impactful Courses on CGPA - ANT Program

Table 4.6: Regression and Correlation Results - ANT Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
ANT321	0.1041	0.4085	<0.001	118
ANT301	0.0101	0.3855	<0.001	129
ANT320	-0.0258	0.3798	<0.001	108
ANT302	0.0630	0.3786	<0.001	101
ANT201	0.0681	0.3588	<0.001	129
ANT420	0.0960	0.3730	<0.001	102
ANT370	0.0093	0.3751	<0.001	95
ANT350	0.0770	0.3720	<0.001	86
ANT104	-0.0166	0.3358	<0.001	111
ANT330	0.0331	0.3442	<0.001	88

These relationships reflect associations with CGPA-which already includes these course grades-and should not be interpreted as causal effects.

(b) APE Program

Figure 4.8 presents the top 10 impactful courses for the APE (Applied Physics and Electronics) program. The regression model yields an R^2 score of **0.4230**, reflecting strong predictive ability. Among the courses, CSE110 and STA201 emerge as the most influential based on their regression coefficients and correlation values.
Table 4.7 shows detailed regression and correlation results. CSE110 (Coef = 0.2489, $r = 0.5602$, $p < 0.001$, count = 32) and STA201 (Coef = 0.1620, $r = 0.4141$) exhibit strong associations with CGPA, whereas MAT203 shows a negative regression coef-

ficient, suggesting it may be inversely associated with overall performance despite a positive correlation.

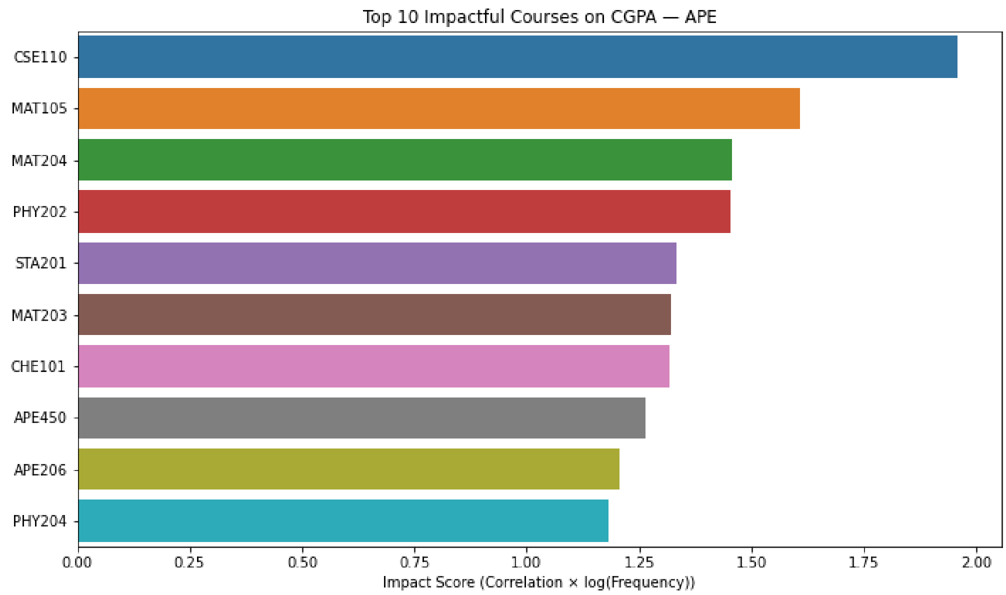

Figure 4.8: Top 10 Impactful Courses on CGPA - APE Program

Table 4.7: Regression and Correlation Results - APE Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
CSE110	0.2489	0.5602	<0.001	32
MAT105	0.0237	0.4729	<0.001	29
MAT204	0.0858	0.4469	<0.001	25
PHY202	0.0207	0.4634	<0.001	22
STA201	0.1620	0.4141	<0.001	24
MAT203	-0.1075	0.4213	<0.001	22
CHE101	0.0904	0.4265	<0.001	21
APE450	0.0861	0.4560	<0.001	15
APE206	0.0114	0.4171	<0.001	17
PHY204	0.0151	0.3945	<0.001	19

(c) ARC Program

The ARC (Architecture) program's top 10 impactful courses are illustrated in Figure 4.9. ARC202 and ARC412 are among the most impactful, as evidenced by their strong regression coefficients and correlation values. The regression model yields an R^2 score of **0.2452**, indicating moderate explanatory strength.
Table 4.8 consolidates regression and Pearson correlation data. ARC202 has the highest coefficient (0.0914) and the strongest correlation ($r = 0.4161$, $p < 0.001$), followed closely by ARC412 and ARC112. Despite their lower coefficients, courses

like ARC214 and ARC411 still maintain significant correlations, reflecting consistent performance patterns across students.

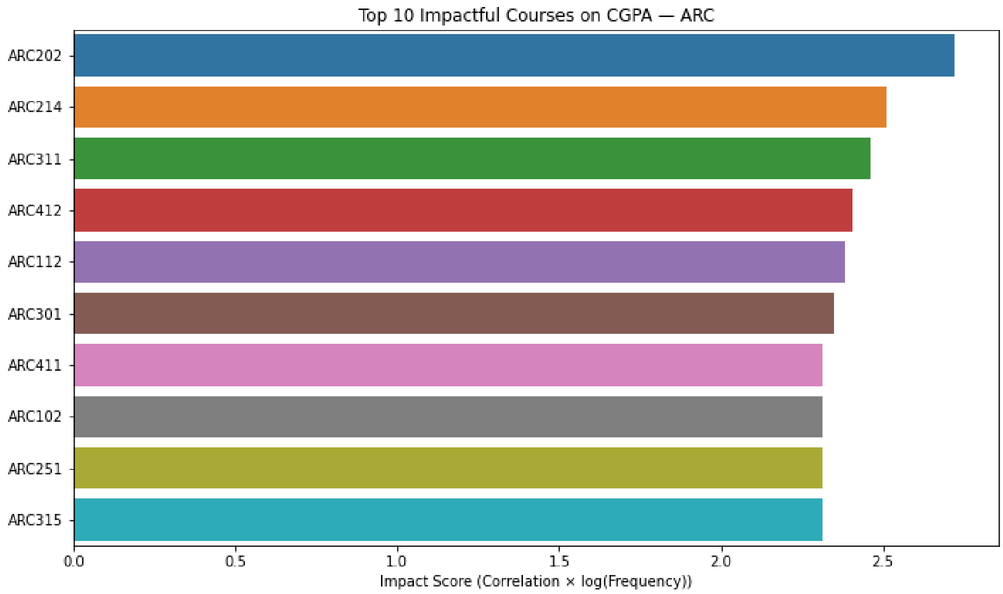

Figure 4.9: Top 10 Impactful Courses on CGPA - ARC Program

Table 4.8: Regression and Correlation Results - ARC Program

Course Code	Coef.	r	p-value	Count
ARC202	0.0914	0.4161	<0.001	691
ARC214	-0.0243	0.3812	<0.001	723
ARC311	0.0609	0.3808	<0.001	645
ARC412	0.0977	0.3753	<0.001	606
ARC112	0.0844	0.3587	<0.001	771
ARC301	0.0824	0.3651	<0.001	622
ARC411	-0.0340	0.3523	<0.001	712
ARC102	0.0447	0.3463	<0.001	795
ARC251	0.0288	0.3515	<0.001	718
ARC315	0.0239	0.3542	<0.001	682

(d) BBA Program

Figure 4.10 highlights the top 10 impactful courses in the BBA (Bachelor of Business Administration) program. The regression model achieves an R^2 score of **0.3472**, indicating strong model fit. MAT101, BUS321, and MGT401 stand out as the most predictive of CGPA.
Table 4.9 contains regression coefficients, correlation values, p-values, student counts, and computed impact scores. MAT101 has the highest coefficient (0.1389) and strong correlation ($r = 0.3919$, $p < 0.001$, count $= 4787$), confirming its central role in shaping academic performance in the BBA curriculum.

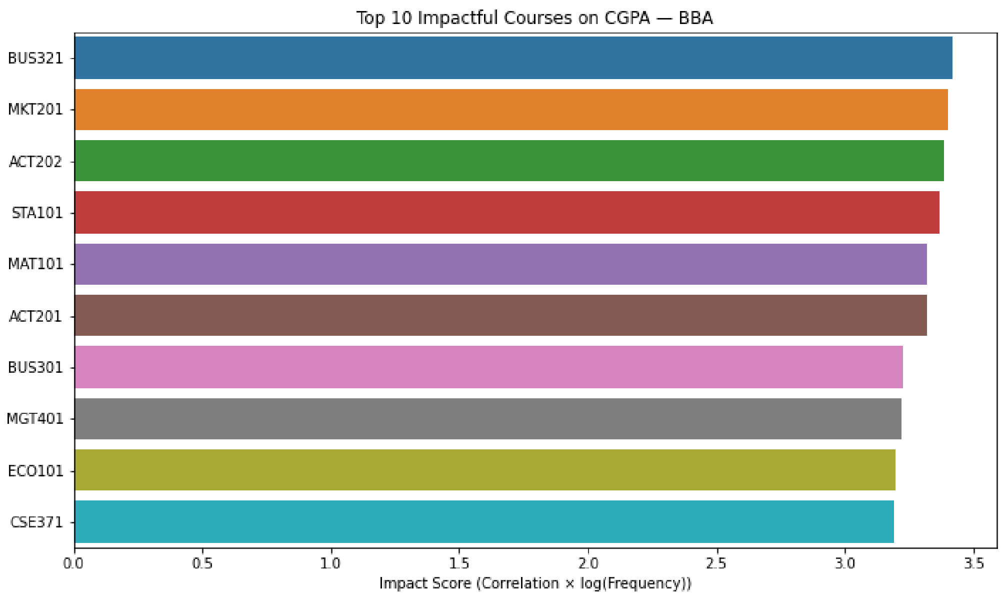

Figure 4.10: Top 10 Impactful Courses on CGPA - BBA Program

Table 4.9: Regression and Correlation Results - BBA Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
BUS321	0.0910	0.4145	<0.001	3822
MKT201	0.0346	0.4046	<0.001	4478
ACT202	0.0431	0.4090	<0.001	3951
STA101	0.0571	0.4015	<0.001	4423
MAT101	0.1389	0.3919	<0.001	4787
ACT201	0.0420	0.3942	<0.001	4506
BUS301	0.0434	0.3924	<0.001	3687
MGT401	0.0671	0.3930	<0.001	3607
ECO101	0.0208	0.3843	<0.001	4101
CSE371	-0.0162	0.3874	<0.001	3761

(e) BIO Program

Figure 4.11 illustrates the top 10 impactful courses in the BIO program. The most influential courses include BTE101, BTE102, and BTE103. The multiple regression model yields an R^2 value of 0.5801, indicating strong explanatory power of the selected features.

Table 4.10 shows that BTE101 has the highest regression coefficient (0.2700) and correlation ($r = 0.6128$), making it the most impactful course. Other courses such as BTE102 and BTE103 also show strong correlations. Courses like BCH201 and BTE202, despite having low or negative coefficients, still exhibit statistically significant correlations, reinforcing their influence.

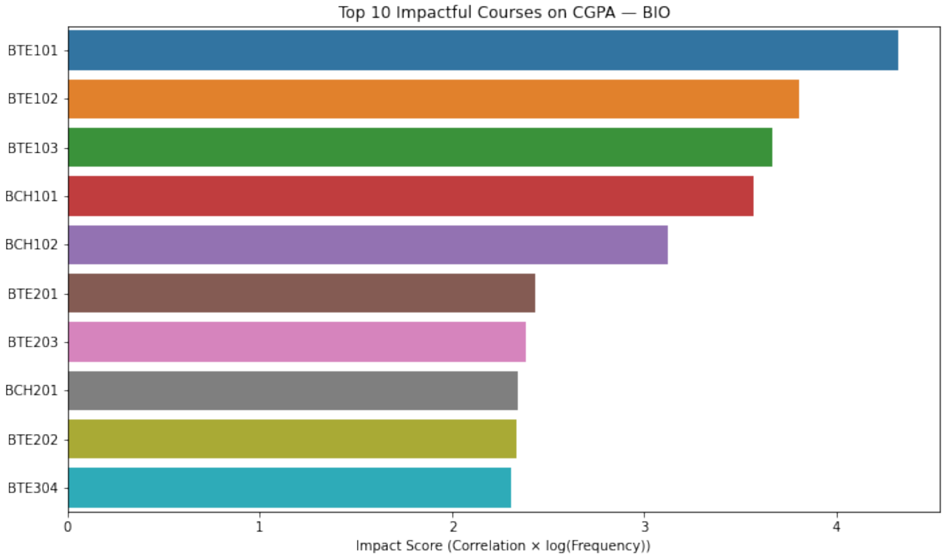

Figure 4.11: Top 10 Impactful Courses on CGPA - BIO Program

Table 4.10: Regression and Correlation Results - BIO Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
BTE101	0.2700	0.6128	<0.001	1155
BTE102	0.1463	0.5507	<0.001	1001
BTE103	0.1346	0.5337	<0.001	963
BCH101	0.0452	0.5134	<0.001	1043
BCH102	0.0031	0.4622	<0.001	858
BTE201	0.0653	0.3834	<0.001	568
BTE203	0.0078	0.3698	<0.001	635
BCH201	-0.0279	0.3603	<0.001	660
BTE202	-0.0021	0.3689	<0.001	558
BTE304	0.0541	0.3709	<0.001	498

(f) CS Program

The top 10 impactful courses for the CS (Computer Science) program are presented in Figure 4.12. The model achieves an R^2 value of 0.5924. Key influential courses include CSE220, CSE111, CSE230, and MAT110.
Table 4.11 reveals that CSE220 has the highest correlation ($r = 0.5828$), followed closely by CSE111 and CSE230. MAT110 and PHY111 also contribute notably, with meaningful coefficients and strong correlations.

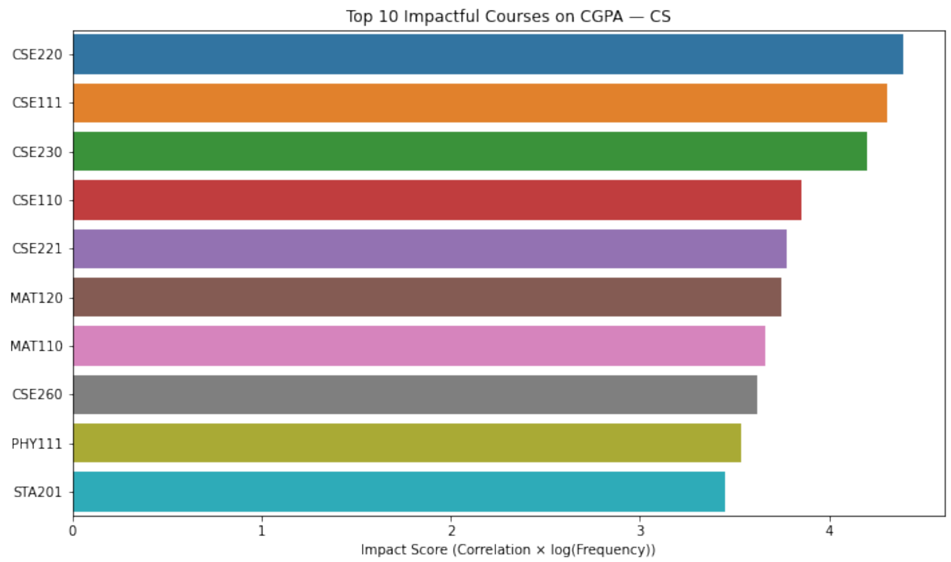

Figure 4.12: Top 10 Impactful Courses on CGPA - CS Program

Table 4.11: Regression and Correlation Results - CS Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
CSE220	0.0993	0.5828	<0.001	1874
CSE111	0.0478	0.5665	<0.001	1995
CSE230	0.0615	0.5570	<0.001	1883
CSE110	0.0790	0.5032	<0.001	2115
CSE221	0.0648	0.5086	<0.001	1667
MAT120	0.0731	0.5037	<0.001	1713
MAT110	0.0966	0.4865	<0.001	1852
CSE260	0.0239	0.4848	<0.001	1749
PHY111	0.0832	0.4729	<0.001	1748
STA201	0.0330	0.4620	<0.001	1758

(g) CSE Program

Figure 4.13 highlights the top courses in the CSE program, with an exceptionally high R^2 value of 0.6549, indicating that the model explains a substantial portion of CGPA variance.

Table 4.12 confirms that CSE110, MAT110, and CSE111 are the most predictive. CSE110 alone shows a regression coefficient of 0.2443 and a very strong correlation ($r = 0.6664$). Courses such as CSE250, CSE220, and PHY111 also show considerable influence.

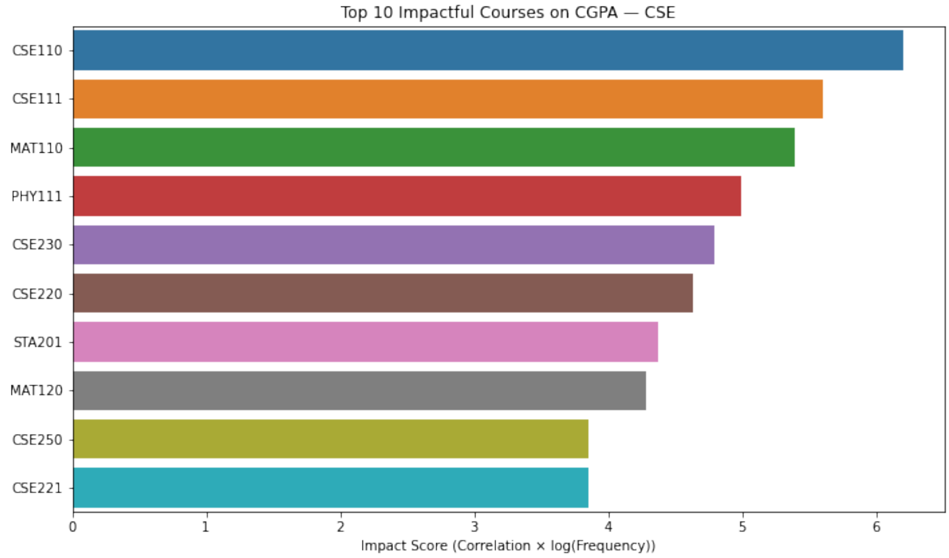

Figure 4.13: Top 10 Impactful Courses on CGPA - CSE Program

Table 4.12: Regression and Correlation Results - CSE Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
CSE110	0.2443	0.6664	<0.001	11086
CSE111	0.0587	0.6143	<0.001	9125
MAT110	0.1959	0.5872	<0.001	9715
PHY111	0.1125	0.5503	<0.001	8772
CSE230	-0.0057	0.5346	<0.001	7882
CSE220	0.0511	0.5243	<0.001	6841
STA201	0.0203	0.4914	<0.001	7285
MAT120	0.0356	0.4848	<0.001	6904
CSE250	0.0516	0.4490	<0.001	5394
CSE221	0.0519	0.4501	<0.001	5232

(h) ECE Program

Figure 4.14 displays the top 10 impactful courses in the ECE program. The regression model explains 20% of the CGPA variance ($R^2 = 0.2000$), which is modest but acceptable for educational datasets.
As seen in Table 4.13, ECE330 has the highest regression coefficient (0.2571), with MAT110 and ACT201 also contributing notably. Although MAT216 and MAT120 have lower coefficients, they remain statistically significant in correlation analysis.

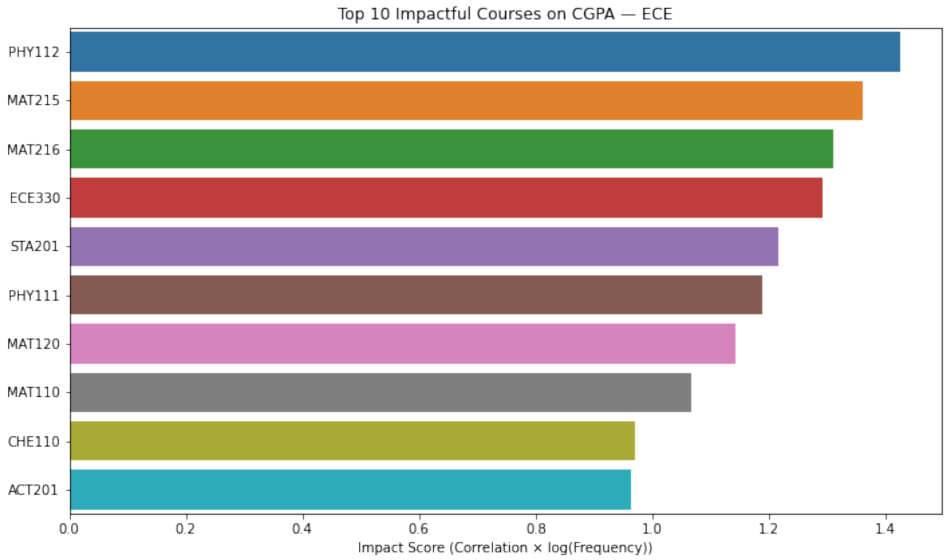

Figure 4.14: Top 10 Impactful Courses on CGPA - ECE Program

Table 4.13: Regression and Correlation Results - ECE Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
PHY112	0.0672	0.2846	<0.001	149
MAT215	0.0388	0.2771	<0.001	135
MAT216	0.0145	0.2717	<0.001	124
ECE330	0.2571	0.2773	<0.001	105
STA201	0.0514	0.2534	<0.001	121
PHY111	0.0515	0.2303	<0.001	173
MAT120	-0.0078	0.2368	<0.001	124
MAT110	0.0843	0.2093	<0.001	162
CHE110	0.0456	0.2204	<0.001	81
ACT201	0.0726	0.2131	<0.001	91

(i) ECO Program

Figure 4.15 presents the top 10 impactful courses for the ECO (Economics) program. The multiple regression model yields an R^2 score of 0.3563, indicating strong explanatory power in predicting CGPA from selected courses. Among the predictors, ECO101 and ECO308 show relatively higher coefficients (0.1311 and 0.1257, respectively), suggesting a strong positive relationship with CGPA.
As shown in Table 4.14, all 10 courses have statistically significant correlations with CGPA ($p < 0.001$), with Pearson r values ranging from 0.4036 (ECO324) to 0.4669 (ECO201). This consistency indicates that students' performances in these courses strongly align with their overall academic success.

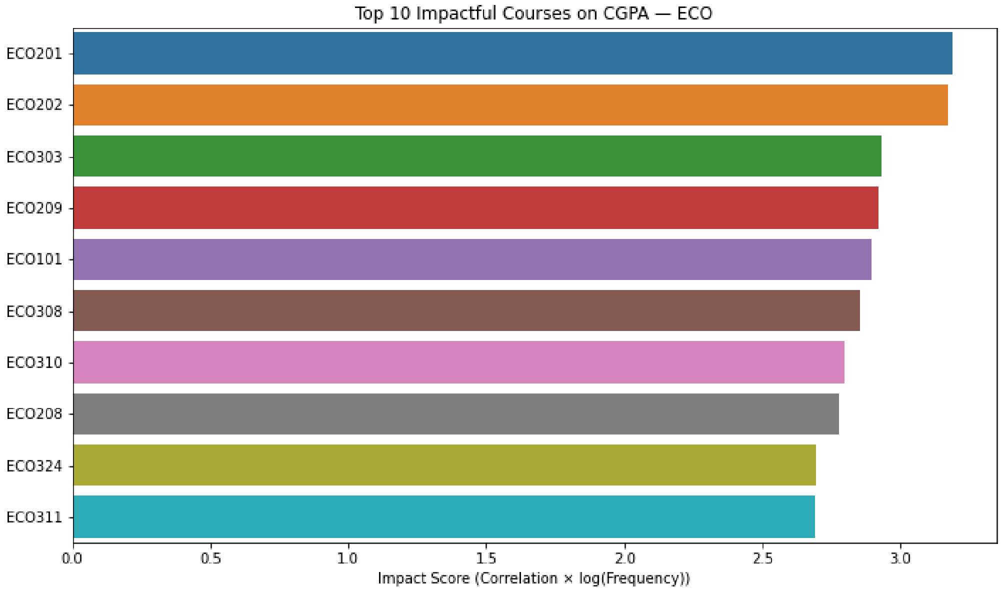

Figure 4.15: Top 10 Impactful Courses on CGPA - ECO Program

Table 4.14: Regression and Correlation Results - ECO Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
ECO201	0.0917	0.4669	<0.001	926
ECO202	0.0518	0.4662	<0.001	902
ECO303	0.0608	0.4437	<0.001	744
ECO209	-0.0056	0.4382	<0.001	788
ECO101	0.1311	0.4168	<0.001	1034
ECO308	0.1257	0.4337	<0.001	719
ECO310	0.0123	0.4179	<0.001	807
ECO208	0.0562	0.4190	<0.001	756
ECO324	0.0258	0.4036	<0.001	797
ECO311	0.0183	0.4098	<0.001	711

(j) EEE Program

Figure 4.16 highlights the top 10 impactful courses for the EEE (Electrical and Electronic Engineering) program. The multiple regression model explains 31.6% of CGPA variance ($R^2 = 0.3160$). EEE241 and CHE110 show the strongest positive regression coefficients, suggesting a tangible link between performance in these courses and overall CGPA.
According to Table 4.15, all courses show significant positive correlations with CGPA ($p < 0.001$). EEE241 stands out with $r = 0.4516$, followed closely by EEE203 ($r = 0.4266$) and EEE305 ($r = 0.4214$). These findings confirm that foundational and advanced core courses significantly influence student academic outcomes in EEE.

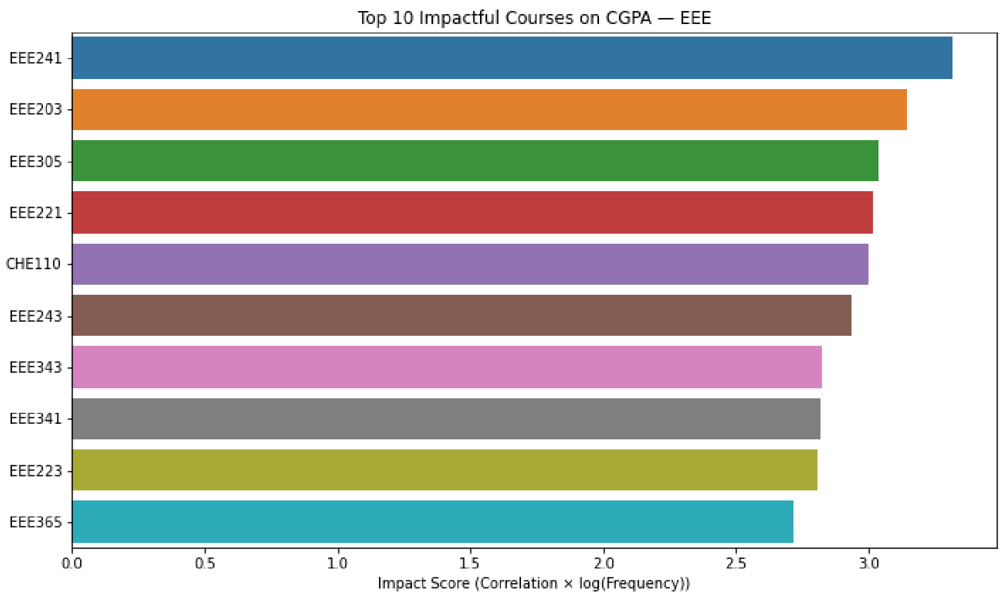

Figure 4.16: Top 10 Impactful Courses on CGPA - EEE Program

Table 4.15: Regression and Correlation Results - EEE Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
EEE241	0.1149	0.4516	<0.001	1542
EEE203	0.0961	0.4266	<0.001	1593
EEE305	0.0745	0.4214	<0.001	1348
EEE221	0.0756	0.4117	<0.001	1506
CHE110	0.0816	0.4042	<0.001	1661
EEE243	0.0003	0.4032	<0.001	1443
EEE343	0.0017	0.3912	<0.001	1368
EEE341	0.0186	0.3913	<0.001	1350
EEE223	0.0714	0.3923	<0.001	1277
EEE365	0.0063	0.3811	<0.001	1244

(k) ENG Program

Figure 4.17 shows the most impactful courses for students in the ENG (English) program. The regression model for this group explains 47.92% of CGPA variance ($R^2 = 0.4792$). ENG111 shows the strongest coefficient (0.1851), reinforcing its importance as a foundational course.
Table 4.16 reveals that all 10 courses have strong, statistically significant correlations with CGPA. Pearson r values exceed 0.47 for all but one course, with ENG111 reaching a maximum r of 0.5444. The high significance of both literature and writing-intensive courses suggests the importance of balanced curriculum design in the English program.

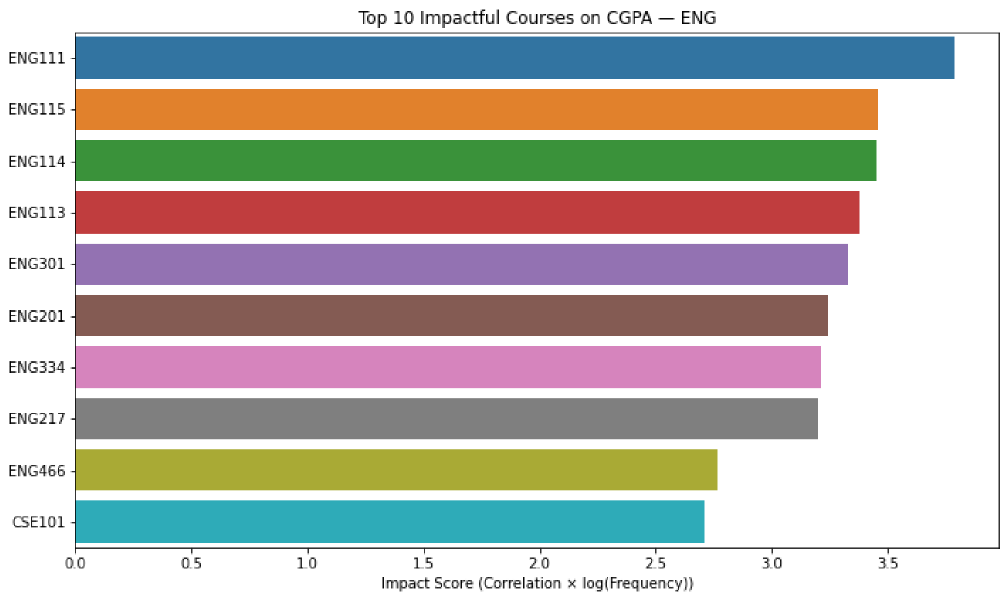

Figure 4.17: Top 10 Impactful Courses on CGPA - ENG Program

Table 4.16: Regression and Correlation Results - ENG Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
ENG111	0.1851	0.5444	<0.001	1053
ENG115	0.0740	0.4968	<0.001	1054
ENG114	0.1085	0.4949	<0.001	1069
ENG113	0.0583	0.4884	<0.001	1004
ENG301	0.0493	0.4938	<0.001	850
ENG201	-0.0320	0.4774	<0.001	888
ENG334	0.0517	0.4823	<0.001	786
ENG217	0.0218	0.4755	<0.001	841
ENG466	0.1228	0.4224	<0.001	697
CSE101	0.0017	0.4107	<0.001	735

(l) LLB Program

Figure 4.18 displays the top 10 impactful law courses for the LLB program. The regression model reports a strong R^2 value of 0.5338. LAW104 (Coef = 0.1639) and LAW102 (0.1021) emerge as the most influential predictors of CGPA.
Table 4.17 confirms these observations, with all correlation values being highly significant ($p < 0.001$). Courses such as LAW304 ($r = 0.4980$) and LAW308 ($r = 0.4817$) also stand out, emphasizing the impact of mid-level and advanced law courses on students' overall academic performance.

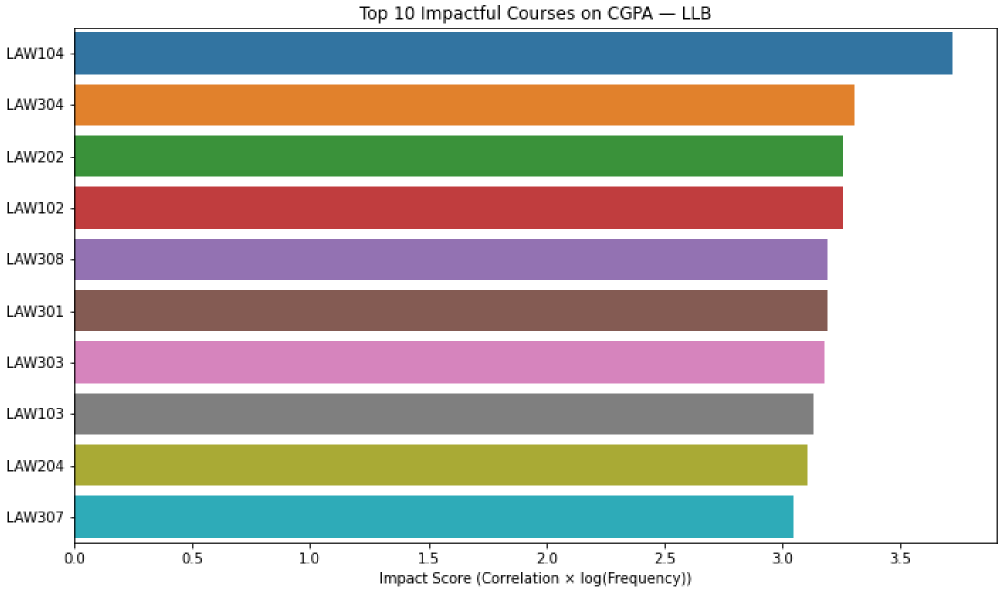

Figure 4.18: Top 10 Impactful Courses on CGPA - LLB Program

Table 4.17: Regression and Correlation Results - LLB Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
LAW104	0.1639	0.5387	<0.001	998
LAW304	0.0553	0.4980	<0.001	766
LAW202	0.0713	0.4734	<0.001	972
LAW102	0.1021	0.4605	<0.001	1172
LAW308	0.0617	0.4817	<0.001	753
LAW301	0.0089	0.4771	<0.001	801
LAW303	0.0270	0.4760	<0.001	791
LAW103	0.0908	0.4501	<0.001	1043
LAW204	0.0487	0.4585	<0.001	877
LAW307	0.0914	0.4657	<0.001	697

(m) MAT Program

Figure 4.19 presents the top 10 impactful courses for the MAT (Mathematics) program. The regression model achieves a strong R^2 of 0.5159, highlighting the predictive relevance of these foundational and advanced courses. Courses like MAT123 (Coef = 0.1853) and MAT311 (0.1622) show the highest coefficients, underscoring their substantial influence on CGPA.

Table 4.18 shows that all selected courses are positively correlated with CGPA, with Pearson r values ranging from 0.3970 to 0.5584. Most p-values are below 0.001, indicating strong statistical significance.

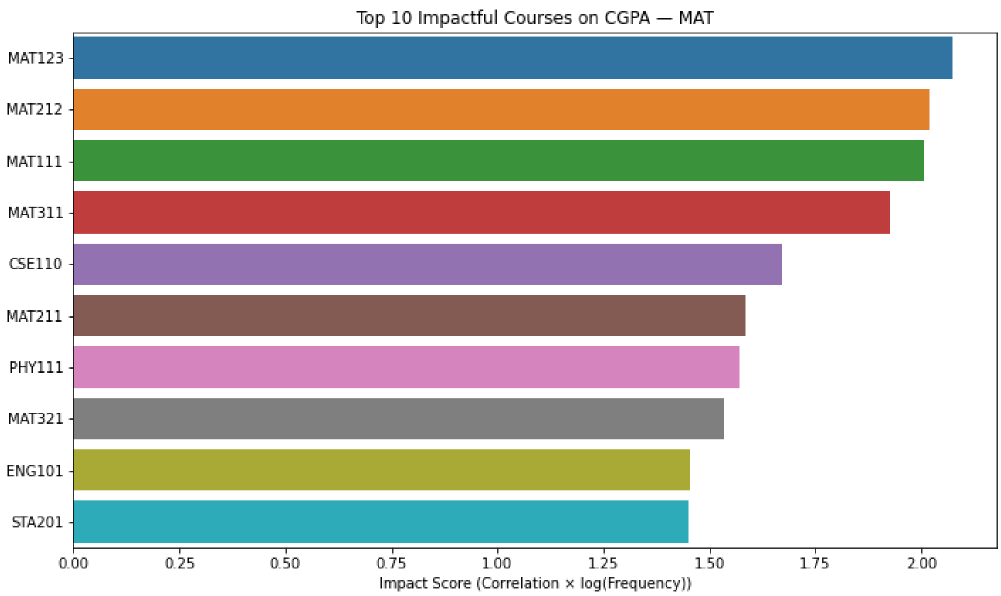

Figure 4.19: Top 10 Impactful Courses on CGPA - MAT Program

Table 4.18: Regression and Correlation Results - MAT Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
MAT123	0.1853	0.5584	<0.001	40
MAT212	0.0641	0.5407	<0.001	41
MAT111	0.0320	0.5104	<0.001	50
MAT311	0.1622	0.5336	<0.001	36
CSE110	0.1015	0.4417	<0.001	43
MAT211	0.1148	0.4489	<0.001	33
PHY111	0.0190	0.4349	<0.001	36
MAT321	-0.0471	0.4466	<0.001	30
ENG101	0.0926	0.3970	<0.001	38
STA201	0.0477	0.4221	<0.001	30

(n) MIC Program

Figure 4.20 illustrates the top 10 impactful courses in the MIC (Microbiology) program. The regression model reports a high R^2 of 0.6328, indicating that these courses jointly explain a large portion of CGPA variation.
Table 4.19 shows that BCH101 (Coef = 0.2213, $r = 0.6256$) and MIC101 (0.1760, $r = 0.6095$) are especially influential. All 10 courses exhibit statistically significant correlations with CGPA ($p < 0.001$), reinforcing their importance in shaping academic performance within the program.

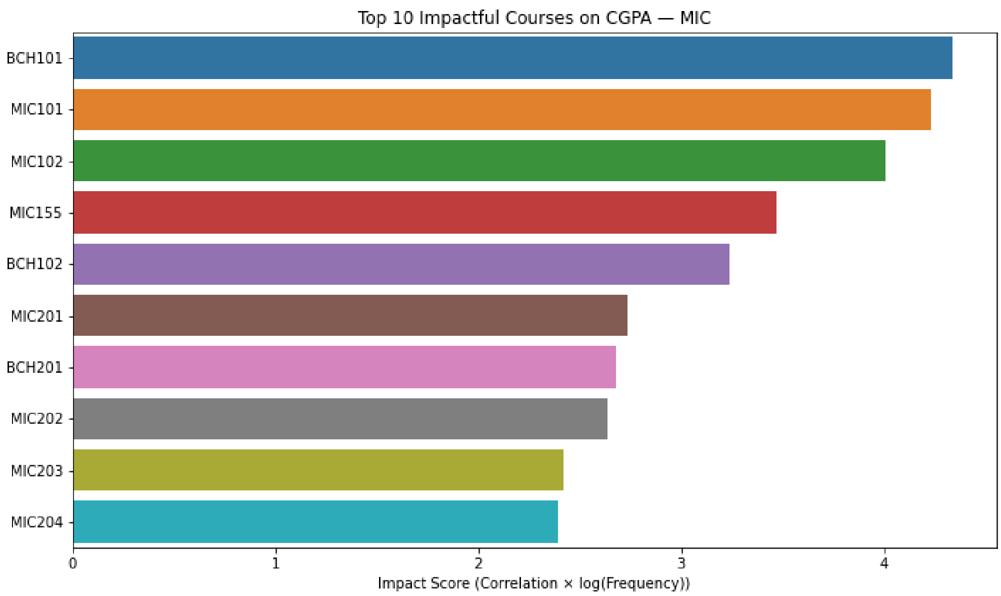

Figure 4.20: Top 10 Impactful Courses on CGPA - MIC Program

Table 4.19: Regression and Correlation Results - MIC Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
BCH101	0.2213	0.6256	<0.001	1026
MIC101	0.1760	0.6095	<0.001	1035
MIC102	0.1414	0.5926	<0.001	869
MIC155	-0.0105	0.5149	<0.001	839
BCH102	0.0785	0.4929	<0.001	713
MIC201	0.0399	0.4252	<0.001	625
BCH201	-0.0016	0.4133	<0.001	651
MIC202	0.0448	0.4158	<0.001	571
MIC203	0.0450	0.3838	<0.001	550
MIC204	0.0133	0.3772	<0.001	565

(o) PHR Program

Figure 4.21 highlights the top 10 impactful courses for the PHR (Pharmacy) program. The regression model achieves an R^2 of 0.4500. PHR111 (Coef = 0.1486) and PHR202 (0.1393) are the top predictors of CGPA.
As shown in Table 4.20, all courses correlate significantly with CGPA ($r > 0.45$, $p < 0.001$). Even courses with negative coefficients like PHR304 (-0.0422) still maintain moderate to strong positive correlation values, suggesting nuanced course effects.

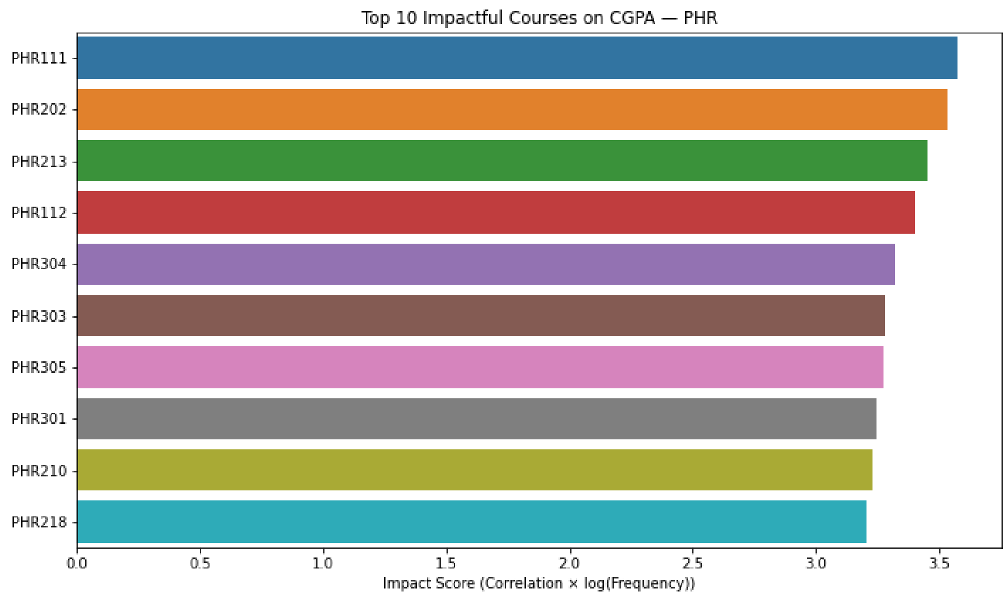

Figure 4.21: Top 10 Impactful Courses on CGPA - PHR Program

Table 4.20: Regression and Correlation Results - PHR Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
PHR111	0.1486	0.5156	<0.001	1026
PHR202	0.1393	0.5107	<0.001	1016
PHR213	0.0786	0.5041	<0.001	943
PHR112	0.0814	0.4862	<0.001	1089
PHR304	-0.0422	0.4881	<0.001	903
PHR303	0.0531	0.4841	<0.001	873
PHR305	0.0958	0.4836	<0.001	872
PHR301	0.0752	0.4802	<0.001	863
PHR210	0.0477	0.4739	<0.001	908
PHR218	-0.1011	0.4540	<0.001	1170

(p) PHY Program

Figure 4.22 displays the most impactful courses in the PHY (Physics) program. The model reports a robust R^2 of 0.4947, with courses like PHY114 (Coef = 0.1890) and PHY110 (0.1766) showing the strongest positive contributions.
As detailed in Table 4.21, most courses are significantly correlated with CGPA. Pearson r ranges from 0.4459 to 0.6203, with all p-values under 0.001. These findings validate that key theoretical and lab-based physics courses significantly influence students' academic performance.

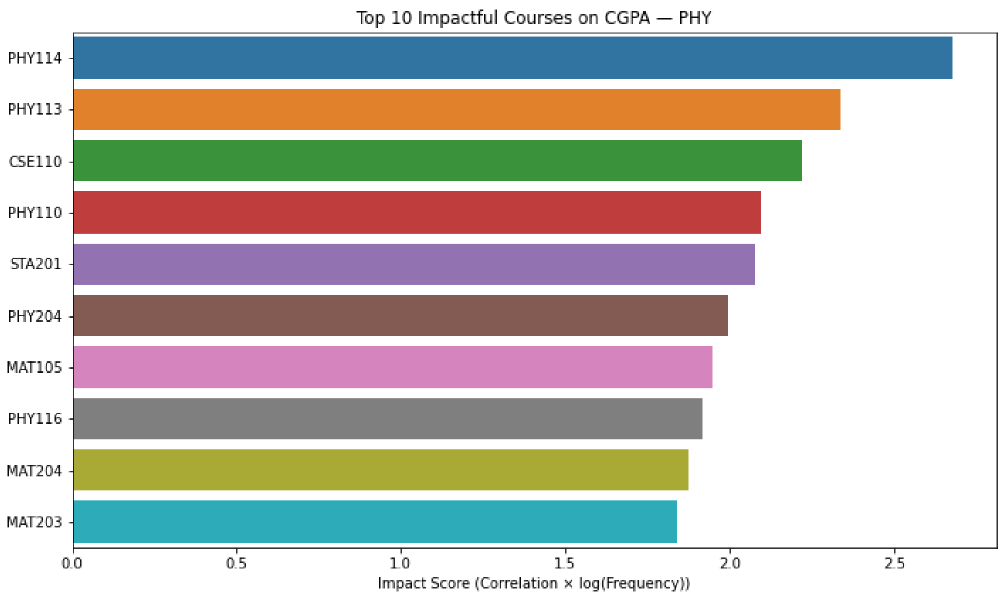

Figure 4.22: Top 10 Impactful Courses on CGPA - PHY Program

Table 4.21: Regression and Correlation Results - PHY Program

Course Code	**Coef.**	**r**	***p*-value**	**Count**
PHY114	0.1890	0.6203	<0.001	74
PHY113	0.0512	0.5554	<0.001	66
CSE110	0.1030	0.5123	<0.001	75
PHY110	0.1766	0.4699	<0.001	85
STA201	0.0163	0.4954	<0.001	65
PHY204	0.0379	0.4915	<0.001	57
MAT105	-0.0142	0.4459	<0.001	78
PHY116	-0.0140	0.4579	<0.001	65
MAT204	0.0463	0.4764	<0.001	50
MAT203	0.0683	0.4675	<0.001	50

4.2.1 Summary of Findings for Hypothesis 2

The findings provide strong support for Hypothesis 2, confirming that certain courses within each academic program hold significantly higher influence on students' cumulative GPA. Key observations include:

- **Consistently High-Impact Courses:** Foundational courses such as `CSE110`, `MAT110`, `PHY111`, and `STA201` appeared as high-impact predictors across multiple programs including CSE, CS, APE, PHY, and MAT. Their high coefficients and strong correlations reinforce their role as academic cornerstones.
- **Program-Specific Key Courses:** Distinctive disciplinary influences emerged–e.g., `ENG111` and `ENG114` in ENG; `LAW104` and `LAW102` in LLB; `PHR111` and

`PHR202` in PHR-highlighting that high-performing students tend to excel in certain core domain courses which serve as performance anchors.

- **Robust Statistical Evidence:** Multiple regression models yielded moderate to high R^2 scores, such as 0.6549 for CSE, 0.5801 for BIO, and 0.5159 for MAT, indicating substantial explanatory power of course-level performance. Corresponding Pearson r values consistently exceeded 0.4, with nearly all p-values below 0.001, confirming strong significance.

- **Variability Across Programs:** While some programs (e.g., MBA, MIC, ENG) exhibit highly predictive course patterns, others like ECE and ANT display more dispersed contributions, indicating differing curriculum structures or grading dynamics.

Overall, this analysis offers a data-driven roadmap for academic planning, highlighting the importance of targeted course support, curricular redesign, and prerequisite alignment. Departments may use these insights to identify bottleneck or gateway courses and allocate resources accordingly to enhance student academic outcomes.

4.3 Hypothesis 3: Prerequisite Impact on Core Course Performance (Department of CSE Only)

Hypothesis: *Prerequisite courses (e.g., ENG091, MAT092) play a critical role in preparing students for success in core programming courses.*
This hypothesis examines the relationship between performance in prerequisite and subsequent core courses, specifically in programming, math, physics, and English sequences. The central question explored is whether a student's success in foundational or prior courses significantly predicts performance in later, more advanced ones.

4.3.1 Exploratory Analysis and Model Results

Correlation and simple linear regression analyses were conducted for 41 course pairings with valid sample sizes. Most course sequences exhibit moderate to strong positive Pearson correlations, with highly significant p-values ($p < 0.001$ in most cases), supporting the hypothesis.
Notably, the strongest correlation ($r = 0.5962$) is observed between **CSE321 and CSE420**, while the highest coefficient of determination ($\mathbf{R^2 = 0.3555}$) is also from the same pair, indicating that over 35% of the variation in CSE420 performance can be explained by performance in CSE321. Similarly, foundational programming sequences such as **CSE110 → CSE111** ($\mathbf{R^2 = 0.2439}$) and **CSE111 → CSE220** ($\mathbf{R^2 = 0.2499}$) demonstrate consistent strength, reinforcing the role of prerequisites. Among math sequences, **MAT216 → MAT215** yields a high correlation of **0.5726**, while physics courses like **PHY112 → CSE250** and **PHY111 → PHY112** also show strong relationships. Interestingly, even the bridge English courses such as **ENG101 → ENG102** exhibit positive correlation ($\mathbf{r = 0.3420}$), suggesting linguistic progression aligns with academic success.

4.3.2 Summary of Findings from Table

Table 4.23 and Figures 4.23-4.24 presents the correlation and regression statistics for 41 prerequisite-to-core course pairs, primarily from the Computer Science and Engineering (CSE), Mathematics (MAT), Physics (PHY), and English (ENG) departments. The table includes the Pearson correlation coefficient (r), corresponding p-values, coefficient of determination (R^2), and regression slope for each course pair. A clear pattern emerges from the data: most pairs exhibit **moderate to strong positive correlations**, with r values typically ranging from 0.40 to 0.60 and $p < 0.001$, indicating high statistical significance within the educational context of this study. These relationships imply that better performance in the prerequisite course is strongly associated with better performance in the corresponding core course. Highlights from the table 4.22 include:

Table 4.22: Summary of Course-to-Course Correlations and Coefficients of Determination (R^2)

Course Pair	r	R^2	**Remarks**
Strongest relationships			
CSE321 → CSE420	0.5962	0.3555	Over 35% of CSE420 variance explained by CSE321.
CSE221 → CSE331	0.5791	0.3354	
CSE370 → CSE110/CSE391	0.5780	0.3350	
MAT216 → MAT325	0.5726	0.3279	
Foundational sequences			
CSE110 → CSE111	0.4938	0.2439	
MAT110 → MAT215	0.4883	0.2384	
PHY111 → PHY112	0.4805	0.2309	
Relatively weaker but significant			
ENG091 → ENG101	0.2956	0.0874	
MAT092 → MAT110	0.2049	0.0420	
CSE311 → CSE410	0.2860	–	$p = 0.057$, not statistically significant at 5% level.

Overall, the data confirms that performance in prerequisite courses is a strong predictor of performance in corresponding core or advanced courses, especially in programming and mathematics-heavy sequences.

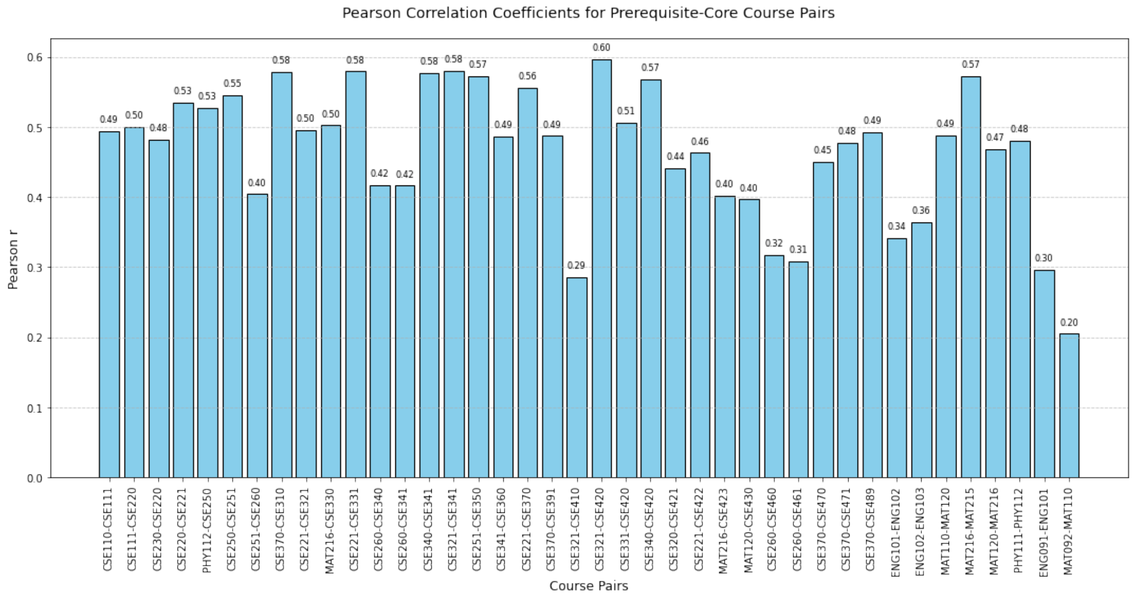

Figure 4.23: Bar Plot: Pearson Correlation Coefficients for Prerequisite-Core Course Pairs

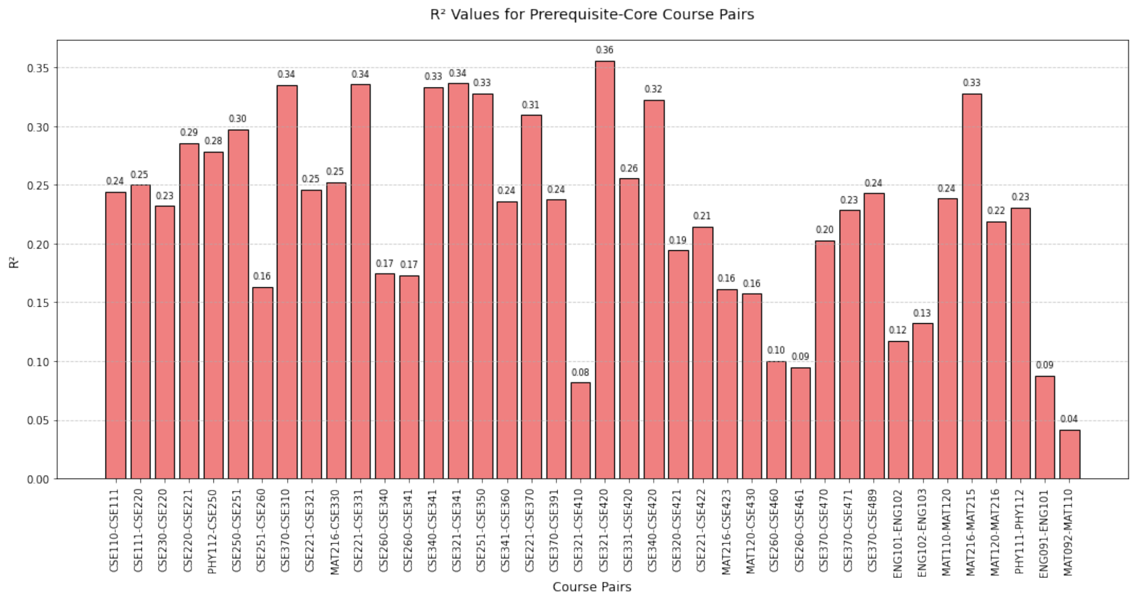

Figure 4.24: Bar Plot: R2 Values for Prerequisite-Core Course Pairs

Table 4.23: Summary of Correlation and Regression Results for Prerequisite-Core Course Pairs

Prereq Course	Next Course	Pearson r	p-value	R^2	Slope
CSE110	CSE111	0.4938	<0.001	0.2439	0.6020
CSE111	CSE220	0.4999	<0.001	0.2499	0.5487
CSE230	CSE220	0.4818	<0.001	0.2321	0.5672
CSE220	CSE221	0.5341	<0.001	0.2853	0.4986
PHY112	CSE250	0.5276	<0.001	0.2784	0.5617
CSE250	CSE251	0.5455	<0.001	0.2975	0.6265
CSE251	CSE260	0.4040	<0.001	0.1632	0.3349
CSE370	CSE310	0.5788	<0.001	0.3351	0.6833
CSE221	CSE321	0.4959	<0.001	0.2459	0.5542
MAT216	CSE330	0.5022	<0.001	0.2522	0.5302
CSE221	CSE331	0.5791	<0.001	0.3354	0.6141
CSE260	CSE340	0.4175	<0.001	0.1743	0.5056
CSE260	CSE341	0.4161	<0.001	0.1731	0.4941
CSE340	CSE341	0.5774	<0.001	0.3333	0.5383
CSE321	CSE341	0.5800	<0.001	0.3364	0.5705
CSE251	CSE350	0.5727	<0.001	0.3280	0.6079
CSE341	CSE360	0.4859	<0.001	0.2361	0.3792
CSE221	CSE370	0.5562	<0.001	0.3093	0.4979
CSE370	CSE391	0.4871	<0.001	0.2373	0.6046
CSE321	CSE410	0.2860	0.057	0.0818	0.1567
CSE321	CSE420	0.5962	<0.001	0.3555	0.5807
CSE331	CSE420	0.5055	<0.001	0.2555	0.5206
CSE340	CSE420	0.5677	<0.001	0.3223	0.5125
CSE320	CSE421	0.4406	<0.001	0.1942	0.5953
CSE221	CSE422	0.4631	<0.001	0.2145	0.5202
MAT216	CSE423	0.4018	<0.001	0.1614	0.4138
MAT120	CSE430	0.3965	<0.001	0.1572	0.2550
CSE260	CSE460	0.3168	<0.001	0.1003	0.4009
CSE260	CSE461	0.3080	<0.001	0.0949	0.2606
CSE370	CSE470	0.4500	<0.001	0.2025	0.4958
CSE370	CSE471	0.4778	<0.001	0.2283	0.4490
CSE370	CSE489	0.4929	<0.001	0.2429	0.5728
ENG101	ENG102	0.3420	<0.001	0.1170	0.3765
ENG102	ENG103	0.3637	0.005	0.1322	0.7455
MAT110	MAT120	0.4883	<0.001	0.2384	0.5322
MAT216	MAT215	0.5726	<0.001	0.3279	0.6010
MAT120	MAT216	0.4680	<0.001	0.2190	0.4502
PHY111	PHY112	0.4805	<0.001	0.2309	0.5245
ENG091	ENG101	0.2956	<0.001	0.0874	0.4350
MAT092	MAT110	0.2049	<0.001	0.0420	0.3782

4.3.3 Summary of Findings for Hypothesis 3

The results strongly validate Hypothesis 3, affirming that prerequisite courses significantly influence student performance in subsequent core and advanced courses. Across 41 analyzed course pairs, consistent patterns of moderate to strong positive correlations, high regression slopes, and statistically significant p-values were observed, with nearly all pairs being significant, except for one marginal case (*CSE321* $\rightarrow$ *CSE410*, $p = 0.057$).
Courses such as **CSE321, CSE370, CSE110, MAT216, and PHY112** act as crucial academic stepping stones, shaping students' capacity to perform well in their follow-up courses like **CSE420, CSE310, CSE111, MAT215, and CSE250**, respectively. In particular, several of the strongest relationships (e.g., $R^2 > 0.33$) suggest that one-third or more of the variance in student grades can be attributed to prerequisite performance.
Interestingly, even early-stage foundational courses like **ENG091** $\rightarrow$ **ENG101** and **MAT092** $\rightarrow$ **MAT110** show statistically significant though comparatively weaker correlations. These findings highlight the long-range impact of academic preparedness and suggest that even introductory remedial courses play a role in a student's academic trajectory.
The evidence supports actionable insights for institutional policy:

- Students struggling in prerequisites should be flagged for additional support (e.g., tutoring, advising) before advancing.
- Academic advisors may use prerequisite performance as a predictive tool for recommending pacing or course load.
- Curriculum committees should review the structure and instructional quality of high-impact prerequisites, as improvements here may yield downstream gains.

In sum, the hypothesis is empirically supported: prerequisite courses are not only foundational in content but foundational in predictive academic value. Strengthening the delivery, assessment, and student support mechanisms in these courses could directly enhance overall academic success across programs.

4.4 Hypothesis 4: Departmental Variation in COVID's Academic Impact

Hypothesis: *The pandemic has a variable impact on student performance across disciplines, with some departments or courses more affected than others.*
This hypothesis investigates whether the COVID-19 pandemic had a uniform impact on student GPA across departments. To address this, we analyze GPA trends and grade distributions using both visual and model-based statistical approaches.

4.4.1 Exploratory Visualizations

Figure 4.25 illustrates the distribution of GPA scores across three COVID periods: Pre-COVID, During-COVID and Post-COVID-segmented by student department.

Each boxplot shows the median (horizontal line), interquartile range (box), and the full range excluding outliers (whiskers), with individual outliers represented as dots. During the COVID period, most departments exhibited a slight upward shift in GPA, with median values clustering closer to the higher end of the 4.0 scale. This was particularly noticeable in departments such as ENH, SoL, and MNS, which demonstrated visibly higher medians and a narrower interquartile range, suggesting both an overall improvement in grades and reduced grade variability. In contrast, departments like CSE and EEE maintained broader distributions, hinting at more diverse academic performance outcomes even during the pandemic.

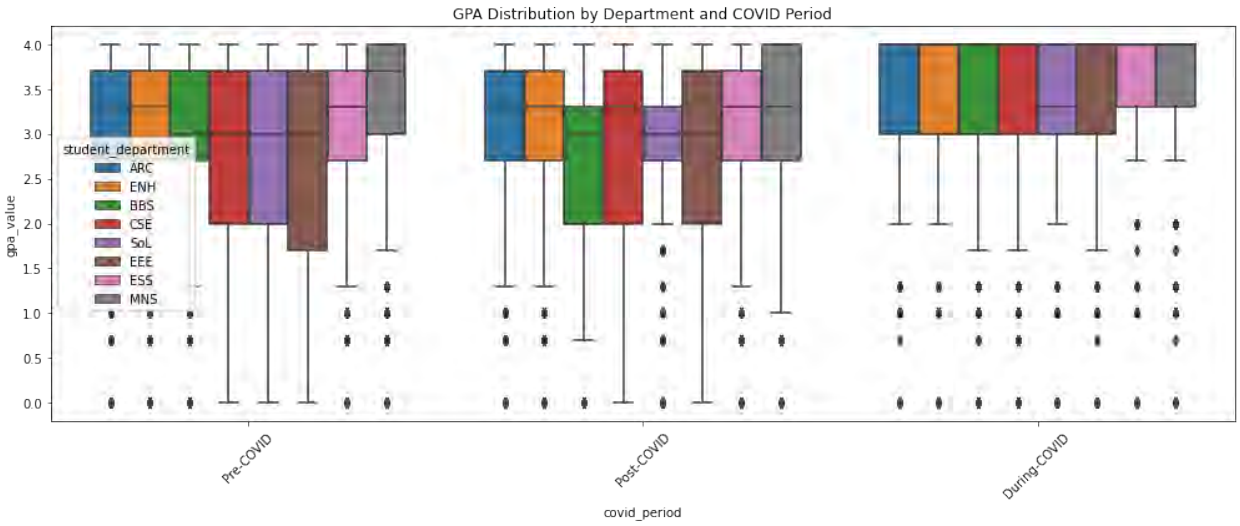

Figure 4.25: Boxplot of GPA across departments by COVID period. Median GPA increases during COVID with more outliers.

In the Post-COVID period, the distribution generally shifted downward in several departments, notably in SoL and CSE, indicating a potential return to stricter grading or normalization of assessment practices after pandemic-related adjustments. Conversely, departments such as MNS and ESS retained relatively higher GPA medians post-COVID, suggesting continued resilience or structural differences in curriculum delivery and evaluation.
Overall, the figure provides a visual confirmation of the differential impact of the pandemic across departments and supports the statistical evidence of interaction effects between COVID period and academic discipline on student performance.

Figure 4.26 presents violin plots showing the distribution and density of GPA scores across departments for each of the three COVID periods: Pre-COVID, During-COVID, and Post-COVID. Unlike boxplots, violin plots combine the features of a boxplot with a kernel density estimate, offering deeper insight into the shape and spread of the data.
During the COVID period, several departments exhibit wider and more symmetric violin shapes with peaks concentrated near the upper end of the GPA scale (e.g., SoL, ENH, and MNS), indicating a higher frequency of students achieving top grades. The density concentration around GPA 3.5-4.0 is noticeably stronger during this period than in Pre-COVID or Post-COVID, suggesting a general inflation of academic performance or easing of grading criteria.

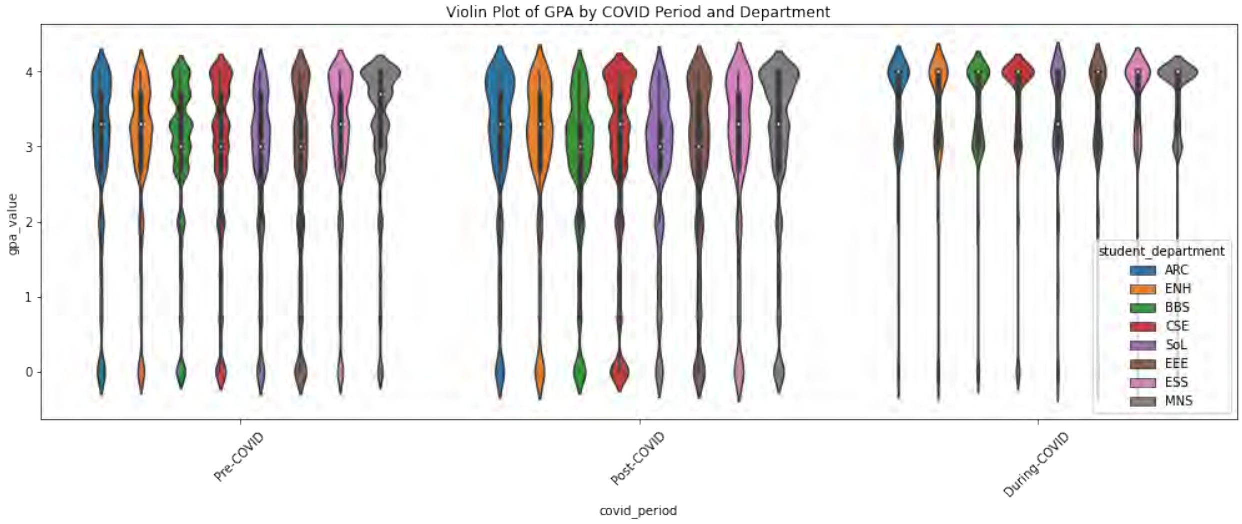

Figure 4.26: Violin plot showing GPA distribution per department during different COVID periods.

In contrast, the Pre-COVID and Post-COVID distributions are more varied and in some cases bimodal or skewed, especially for departments like CSE and EEE. These irregular shapes reflect a broader spread of GPA scores, implying a return to more typical performance variability or stricter evaluation methods outside the COVID period.
The violin plots also highlight department-specific differences. For instance, MNS and ESS consistently demonstrate tight, dense distributions with high GPA medians, while CSE and BBS show broader distributions with more pronounced tails, indicating higher variability and a greater number of students with lower GPAs.
Overall, the figure reinforces the statistical findings that GPA distributions shifted significantly during the pandemic, and that the magnitude and direction of these shifts were not consistent across departments.

Figure 4.27 illustrates the longitudinal trends in average GPA across academic semesters from various departments, spanning Pre-COVID, During-COVID, and Post-COVID periods. Each line represents one department, plotting mean GPA values for each semester as defined by the `year_semester` variable.
Prior to 2020, the GPA trends were relatively stable for most departments, though some variability is observed, particularly in departments like ENH, SoL, and BBS, which showed occasional spikes and dips. Around the onset of the COVID-19 pandemic in Spring 2020 (semester 20201), there is a noticeable upward shift in GPA across nearly all departments. This sharp rise is especially prominent in departments such as SoL, ESS, and MNS, suggesting possible systemic changes in assessment, grading flexibility, or instructional adaptations during remote learning periods.

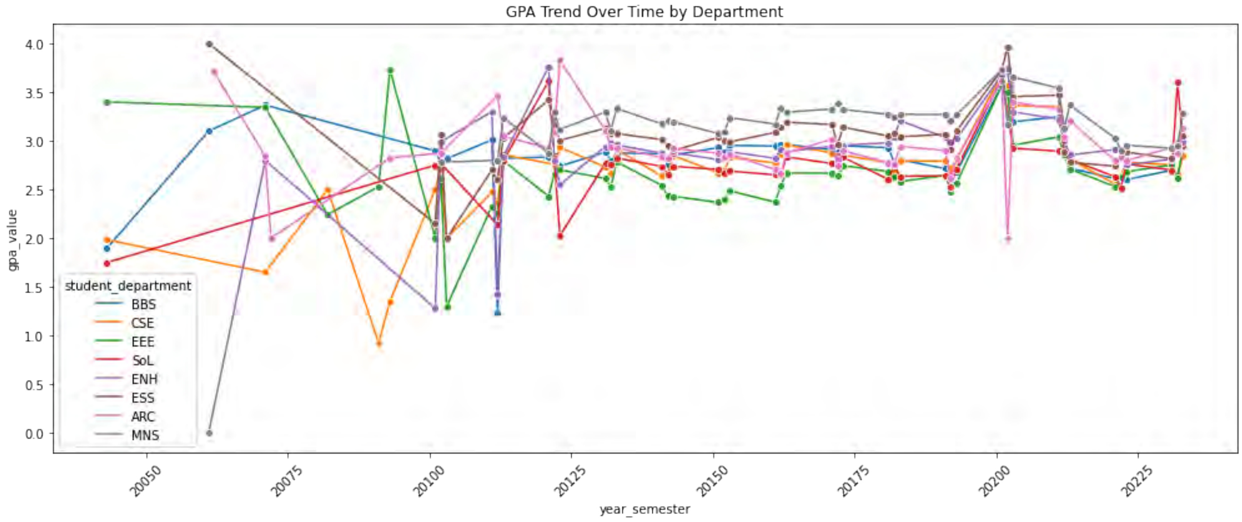

Figure 4.27: Line plot of GPA progression by semester and department. The COVID phase is associated with notable GPA shifts.

Following this peak, the Post-COVID semesters reveal a decline in GPA levels in most departments, indicating a reversion to pre-pandemic grading patterns. Departments such as CSE and BBS, in particular, experienced visible drops, suggesting that any academic leniencies or structural changes that temporarily benefited student performance may have been rolled back.
The plot also shows that some departments, such as MNS and ESS, maintained consistently higher GPA averages throughout the timeline, which may reflect curriculum structure, departmental policies, or student cohort characteristics. Conversely, CSE and EEE demonstrate relatively lower GPA trajectories across semesters.
Overall, this figure visually confirms the temporal fluctuations in academic performance surrounding the COVID-19 pandemic and underscores the non-uniformity of its impact across different academic disciplines.

Figure 4.28 presents a grouped bar chart showing the average GPA across departments for each of the three COVID periods: Pre-COVID, During-COVID, and Post-COVID. This visualization offers a concise comparative snapshot of how departmental GPA levels shifted over time, making it easier to identify both general trends and department-specific deviations.
The During-COVID period exhibits the highest GPA levels across all departments. Departments such as MNS, ESS, and SoL report averages exceeding 3.5, while others like ARC, ENH, and CSE also show elevated GPAs compared to the Pre- and Post-COVID periods. This consistent rise supports the broader trend observed in other visualizations, suggesting increased academic leniency, adapted grading practices, or reduced performance variability during the remote learning phase.
In the Post-COVID period, average GPAs declined for most departments, with the largest drops observed in CSE, BBS, and ARC. This likely reflects the reinstatement of pre-pandemic academic standards or changes in assessment practices. Notably, even in the Post-COVID period, departments like MNS and ESS maintained higher average GPA values, indicating resilience or structural stability in academic delivery.
Comparing Pre-COVID and Post-COVID values reveals that while GPA levels largely returned to pre-pandemic baselines, some departments (e.g., SoL and ENH) retained

modest gains, hinting at potential long-term effects of the pandemic on grading or instructional methods.
Overall, this figure reinforces the conclusion that GPA outcomes during and after the pandemic varied significantly by department, supporting the hypothesis that COVID-19 had a discipline-specific impact on academic performance.

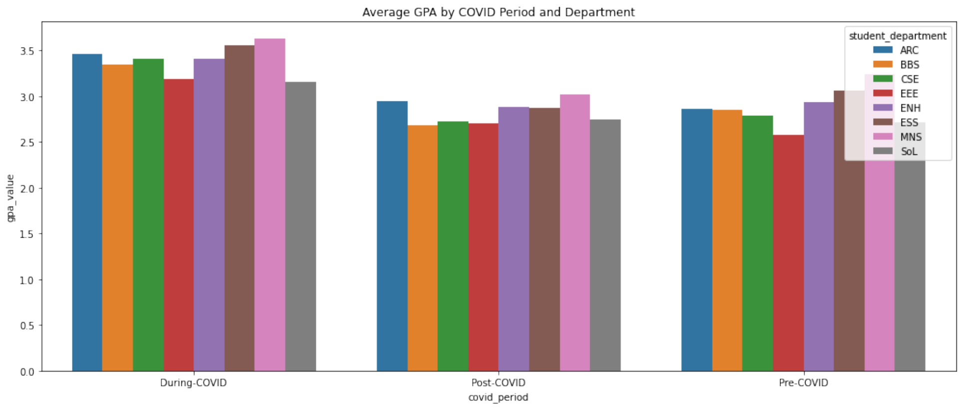

Figure 4.28: Average GPA by department and COVID period. MNS and ESS show higher stability.

4.4.2 Statistical Model Results

(a) Two-Way ANOVA

Table 4.24 confirms statistically significant effects of both department and COVID period, as well as their interaction, on GPA. It presents the results of the two-way ANOVA, which examines the independent and interactive effects of the *student_department* and *covid_period* variables on student GPA. The high F-values for both the main effects (Department, $F = 1856.07$ and COVID Period, $F = 12482.92$) indicate statistically significant differences in GPA across departments and time periods. The interaction effect is also significant ($F = 116.34$), suggesting that the impact of the pandemic on GPA varied across different departments. The residual sum of squares captures the remaining unexplained variance in the model. These results collectively support the hypothesis that COVID-19 had a non-uniform impact on academic performance across disciplines.

Table 4.24: Two-way ANOVA results: Significant department, COVID period, and interaction effects; reference categories: Department = ARC (Architecture) ($p < 0.001$)

Source	**Sum of Squares**	**df**	**F-value**
Department	18801.5	7	1856.07
COVID Period	34743.8	2	12482.92
Dept × COVID	2266.7	14	116.34
Residual	1148768.0	825470	–

(b) Multiple Linear Regression

OLS results in Table 4.25 reaffirm the significant impact of both predictors. CSE and SoL had lower GPAs; MNS showed higher performance. Table 4.25 summarizes the output of the multiple linear regression model, which assesses the main effects of selected departments and COVID periods on student GPA. The intercept represents the baseline GPA, with all categorical variables coded relative to a reference group (departments and the "During-COVID" period, which are not shown in the table). Negative coefficients for the CSE (-0.106) and SoL (-0.134) departments indicate that students in these programs had significantly lower average GPA compared to the reference group, while the MNS department had a positive association (+0.248), suggesting comparatively higher GPA. It is noted to be that SoL's negative coefficient in the regression tables is relative to a high-performing reference department (ARC); the visual trend still shows SoL improving during COVID
Furthermore, both the Pre-COVID and Post-COVID periods are associated with significantly lower GPAs relative to the COVID period, with coefficients of -0.639 and -0.630, respectively. All coefficients are statistically significant at the $p < 0.001$ level, confirming that both departmental affiliation and the time period significantly influenced academic performance.

Table 4.25: OLS regression results for GPA by department and COVID period. All listed predictors are statistically significant at the 0.001 level.

Predictor	**Coef.**	**Std. Err**	**t-value**	***p*-value**
Intercept	3.465	0.007	531.4	<0.001
CSE	-0.106	0.006	-17.7	<0.001
SOL	-0.134	0.008	-22.5	<0.001
MNS	0.248	0.007	37.8	<0.001
Post-COVID	-0.630	0.004	-152.8	<0.001
Pre-COVID	-0.639	0.004	-165.7	<0.001

(c) Mixed Effects Model

The mixed-effects model accounts for repeated measures per student. Results Table 4.26 remain consistent with OLS, showing robust predictors and non-trivial student-level variation (Group Var = 0.791). Table 4.26 presents the results of the mixed effects model, which accounts for within-student variance by including random intercepts for each student. The intercept value of 3.388 represents the estimated average GPA for the reference group (students from the omitted department during the "During-COVID" period).
Among the departmental predictors, CSE (-0.144), EEE (-0.286), ENH (-0.090), and SoL (-0.245) are all associated with significantly lower GPAs compared to the reference group, with p-values below 0.05. On the other hand, MNS (+0.180) shows a statistically significant positive effect on GPA, while the effects of BBS and ESS are not statistically significant ($p = 0.536$ and $p = 0.081$, respectively).
The Pre-COVID and Post-COVID periods are both associated with substantial declines in GPA relative to the During-COVID period, with coefficients of -0.639 and -0.592, respectively, both significant at the $p < 0.001$ level.

The estimated group-level variance of 0.791 indicates meaningful variation in GPA across individual students, further justifying the inclusion of random effects in the model.

Table 4.26: Mixed Effects Model results with student-level random intercepts. Both department and COVID period are significant.

Predictor	**Coef.**	**Std. Err**	**z-value**	***p*-value**
Intercept	3.388	0.026	130.187	<0.001
BBS	-0.017	0.028	-0.610	0.536
CSE	-0.144	0.027	-5.313	<0.001
EEE	-0.286	0.030	-9.426	<0.001
ENH	-0.090	0.035	-2.590	0.010
ESS	0.060	0.034	1.747	0.081
MNS	0.180	0.029	6.118	<0.001
SOL	-0.245	0.035	-7.093	<0.001
Post-COVID	-0.592	0.024	-150.842	<0.001
Pre-COVID	-0.639	0.024	-165.680	<0.001
Group Var (student) = 0.791				

(d) GEE Regression

Table 4.27 presents the Generalized Estimating Equations (GEE) output, which also confirms significant fixed effects and supports previous findings under a robust correlation structure. Table 4.27 shows the results from the Generalized Estimating Equation (GEE) regression model, which accounts for within-student correlation using an exchangeable covariance structure. The intercept value of 3.3958 represents the average GPA of students in the reference group - those from the omitted department during the "During-COVID" period.
Similar to the mixed effects model, several departmental predictors are statistically significant. CSE (-0.1409), EEE (-0.2396), ENH (-0.0810), and SoL (-0.2385) are all associated with significantly lower GPAs compared to the reference group, with $p < 0.05$. MNS shows a statistically significant positive coefficient ($+0.1850$), indicating that students in this department performed better on average. BBS and ESS have non-significant coefficients ($p = 0.269$ and $p = 0.052$, respectively), suggesting their effects are not distinguishable from zero at the conventional significance threshold.
Regarding the COVID period, both Pre-COVID (-0.6362) and Post-COVID (-0.5932) are associated with significantly lower GPA compared to the During-COVID period, with very large absolute z-values and p-values well below 0.001. These results reinforce the observed trend that student GPA increased temporarily during the pandemic and declined afterward.

Table 4.27: GEE Regression results with exchangeable covariance structure. COVID periods and most department indicators are significant.

Predictor	**Coef.**	**Std. Err**	**z-value**	***p*-value**
Intercept	3.3958	0.024	140.154	<0.001
BBS	-0.0282	0.026	-1.105	0.269
CSE	-0.1409	0.025	-5.577	<0.001
EEE	-0.2396	0.029	-8.239	<0.001
ENH	-0.0810	0.033	-2.443	0.015
ESS	0.0608	0.032	1.945	0.052
MNS	0.1850	0.027	6.784	<0.001
SOL	-0.2385	0.032	-7.491	<0.001
Post-COVID	-0.5932	0.007	-85.184	<0.001
Pre-COVID	-0.6362	0.007	-90.715	<0.001

(e) **Kruskal-Wallis Test**

As shown in Table 4.28, all departments demonstrated significant GPA shifts across COVID periods, reaffirming non-parametric trends observed visually. Table 4.28 displays the results of the Kruskal-Wallis test, a non-parametric method used to assess whether GPA distributions differ significantly across COVID periods for each department. This test is appropriate when the assumptions of normality or homogeneity of variance are violated.
All departments included in the analysis - CSE, SoL, MNS, BBS, ENH, and ARC - exhibit statistically significant results, with p-values less than 0.001 in each case. The high H-statistics for departments like CSE (23927.06) and SoL (11871.43) indicate substantial differences in GPA distributions across the three COVID periods (Pre-COVID, During-COVID, and Post-COVID). These findings reinforce the hypothesis that the pandemic's academic impact was not consistent across departments, further validating the results obtained from parametric models.

Table 4.28: Kruskal-Wallis test results: GPA distributions significantly differ across COVID periods.

Department	**H-statistic**	***p*-value**
CSE	23927.06	<0.001
SOL	11871.43	<0.001
MNS	6519.27	<0.001
BBS	6442.54	<0.001
ENH	1582.36	<0.001
ARC	1754.73	<0.001

(f) Chi-Square Test of Grade Distributions

Letter grade distributions changed during COVID. Figure 4.29 shows an increase in A and A+ grades during pandemic semesters. The chi-square test result ($\chi^2 = 111577.74$, $p < 0.001$) confirms this shift. Table 4.29 reports the results of the Chi-Square test, which was conducted to determine whether there was a statistically significant difference in the distribution of letter grades across the three COVID periods (Pre-COVID, During-COVID, and Post-COVID). The test yielded a Chi-square statistic of 111577.74 with 22 degrees of freedom, and a p-value less than 0.001.

This extremely low p-value indicates strong evidence against the null hypothesis, suggesting that the distribution of grades was not independent of the COVID period. In other words, grading patterns significantly changed across time, which may reflect shifts in assessment styles, grading policies, or student behavior during the pandemic. These results are also visually supported by the corresponding grade heatmap.

Table 4.29: Chi-Square test results for grade distribution across COVID periods.

Statistic	**Value**
Chi-square Statistic	111577.74
p-value	<0.001
Degrees of Freedom	22

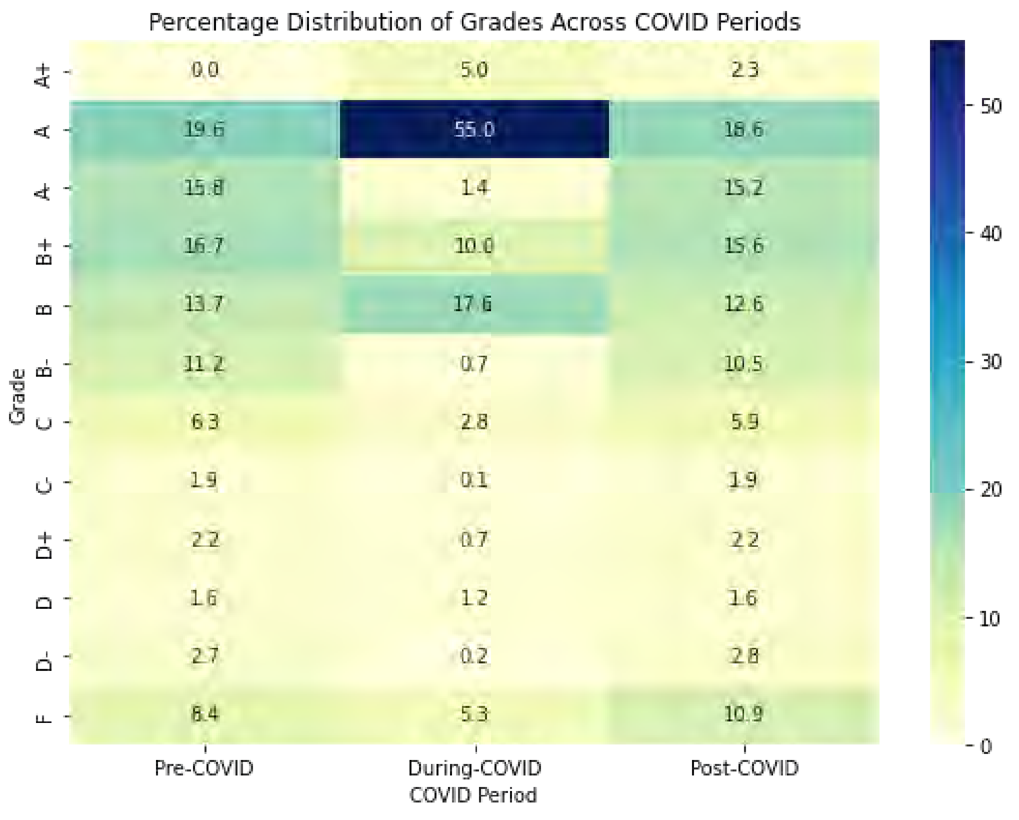

Figure 4.29: Grade distribution heatmap by COVID period. A significant increase in higher grades is observed during COVID.

4.4.3 Summary of Findings for Hypothesis 4

The findings provide strong support for Hypothesis 4, confirming that the academic impact of COVID-19 was not uniform across departments. Key observations are as follows:

- **GPA Trends:** GPA increased significantly during the COVID period across most departments, particularly in ENH and SoL. However, post-COVID semesters showed a decline in GPA in departments such as CSE and SoL, suggesting a possible return to more rigorous assessment standards or reduced leniency.
- **Statistical Significance:** All regression models (OLS, GEE, and Mixed Effects) identified both department and COVID period as significant predictors of GPA. These results were robust across modeling assumptions and confirmed the interaction between discipline and time period.
- **Interaction Effect:** The Two-Way ANOVA revealed not only main effects for department and COVID period, but also a significant interaction effect, indicating that departments responded differently to the challenges posed by the pandemic.
- **Non-Parametric Validation:** The Kruskal-Wallis test results supported the findings from parametric models, showing statistically significant GPA distribution differences across COVID periods within each department.
- **Grade Distribution Shift:** The Chi-Square test revealed significant shifts in grade distributions, with a notable increase in top grades during the COVID semesters, suggesting possible grade inflation or assessment policy adjustments.

Overall, the results confirm that the pandemic's impact on academic performance was heterogeneous across academic disciplines. This reflects a multifaceted response to the crisis, shaped by departmental policies, teaching methods, and student adaptability during and after the pandemic.

4.5 Hypothesis 5: Temporal Impact of COVID-19 on Academic Performance

Hypothesis: *Student academic performance varies significantly across different phases of the COVID-19 pandemic (Pre-COVID, During-COVID, Post-COVID), indicating temporal shifts in GPA outcomes.*
This hypothesis explores the overall temporal impact of the COVID-19 pandemic on student GPA across three distinct periods-Pre-COVID, During-COVID, and Post-COVID-without segmenting by program or department. Both exploratory visualizations and statistical analyses are employed to evaluate GPA trends and grade distributions over time.

4.5.1 Exploratory Visualizations

Figure 4.30 presents a bar chart comparing the average GPA across the three COVID periods. The During-COVID period shows the highest mean GPA at approximately **3.4**, a significant increase from both Pre-COVID (**2.85**) and Post-COVID (**2.8**). This clear elevation in GPA during the pandemic suggests systemic academic leniency or adjustments in grading and assessment styles during remote learning. The decline in GPA post-pandemic indicates a reversion toward more traditional grading standards.

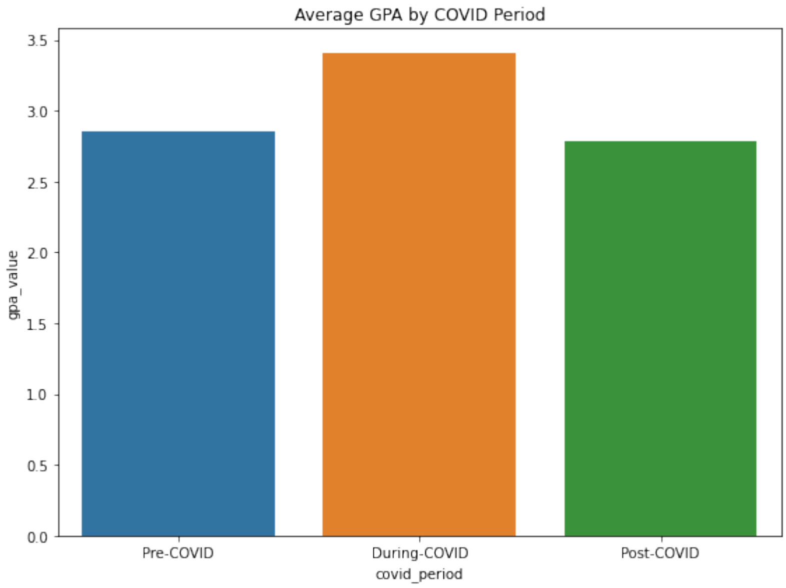

Figure 4.30: Average GPA by COVID period. The highest GPA is observed During-COVID.

Figure 4.31 depicts a line plot illustrating GPA progression over multiple semesters, segmented by academic year and semester. The GPA trend remains relatively stable around **2.7–3.0** for most semesters prior to 2020. However, there is a sharp upward spike in 2020, coinciding with the transition to online learning during the pandemic, peaking at nearly **3.7**. This spike gradually declines in subsequent semesters, returning closer to pre-pandemic levels. The figure provides strong visual evidence of GPA inflation during the pandemic and subsequent normalization.

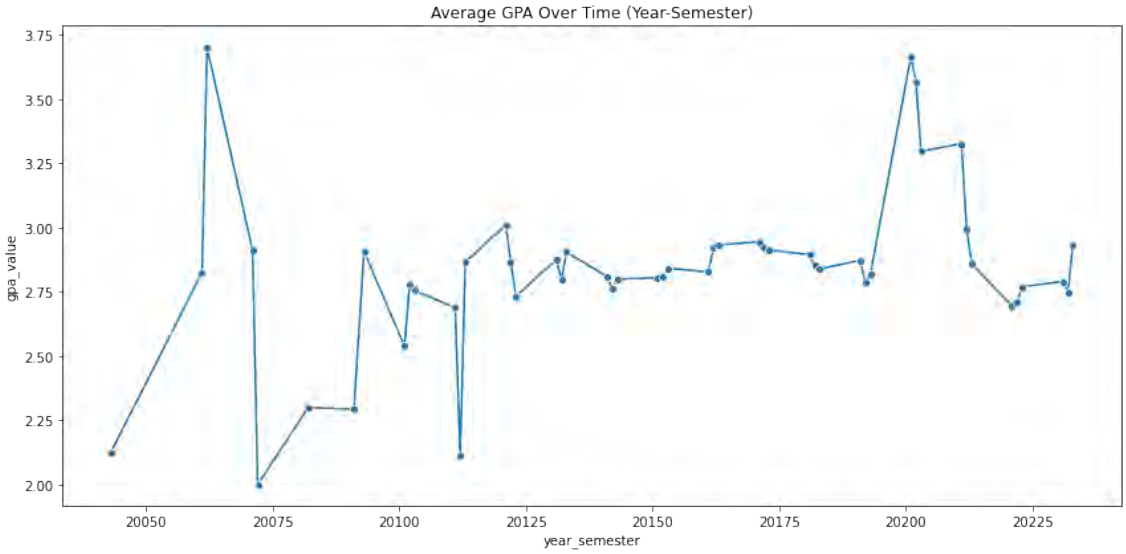

Figure 4.31: GPA trend over time by semester. A spike is observed During-COVID.

Figure 4.32 showcases boxplots comparing GPA distributions across the three pandemic periods. During-COVID distributions exhibit higher medians and reduced variability, with the median GPA around **3.4**, compared to **3.2** Pre-COVID and Post-COVID. Whiskers and outliers are more spread in Pre- and Post-COVID periods, suggesting greater variability and stricter grading. The clustering of GPAs during the pandemic reflects a more uniform grading pattern likely influenced by relaxed academic policies.

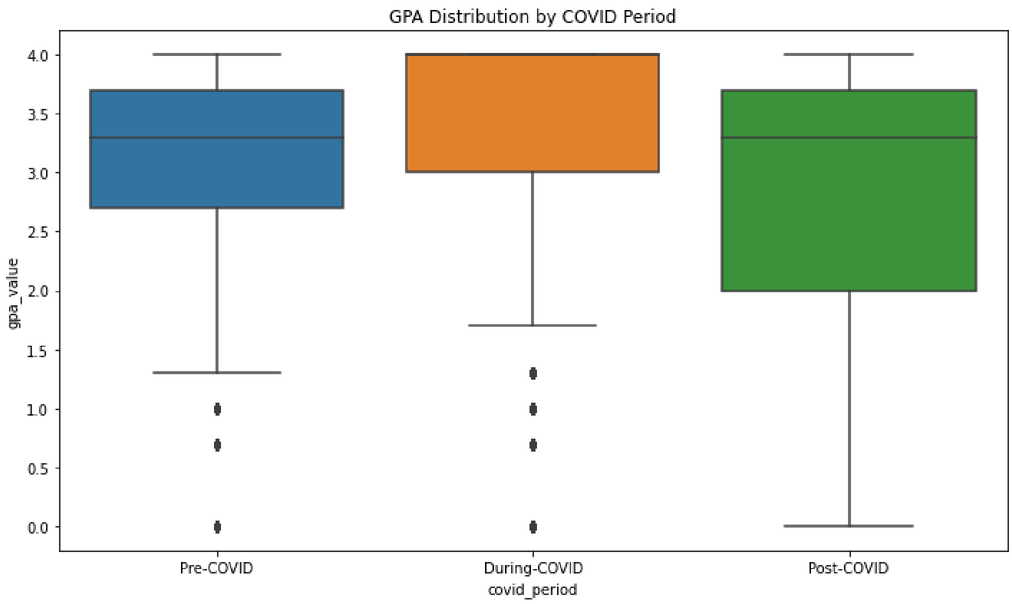

Figure 4.32: Boxplot of GPA distributions by COVID period. Median GPA increases During-COVID.

Figure 4.33 presents a heatmap detailing the percentage distribution of letter grades across COVID periods. The proportion of **A** grades surged from **19.6%** Pre-COVID to **55.0%** During-COVID, alongside a rise in **A+** grades from **0.0%** to **5.0%**. Simultaneously, lower grades such as **B**, **C**, and **F** decreased in frequency, demonstrating a significant shift toward higher grades during the pandemic.

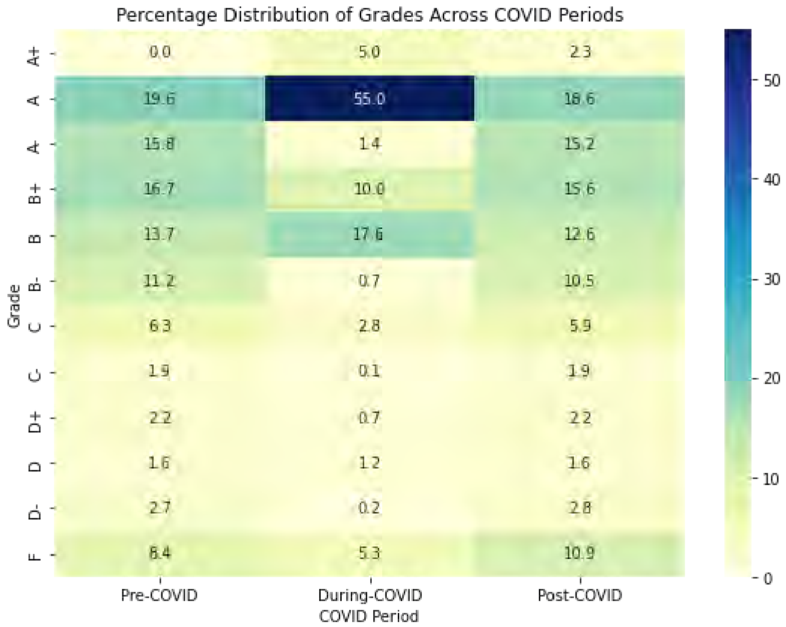

Figure 4.33: Percentage distribution of grades across COVID periods. A and A+ grades increase During-COVID.

Figure 4.34 displays violin plots combining GPA distributions with density estimates across the three COVID periods. The During-COVID violin is narrower and sharply peaked around **3.4–4.0**, indicating a concentration of higher GPA values. Pre-COVID and Post-COVID violins are broader, showing more variability and wider GPA ranges. This visualization underscores the uniform GPA increase during the pandemic and a return to normal variability afterward.

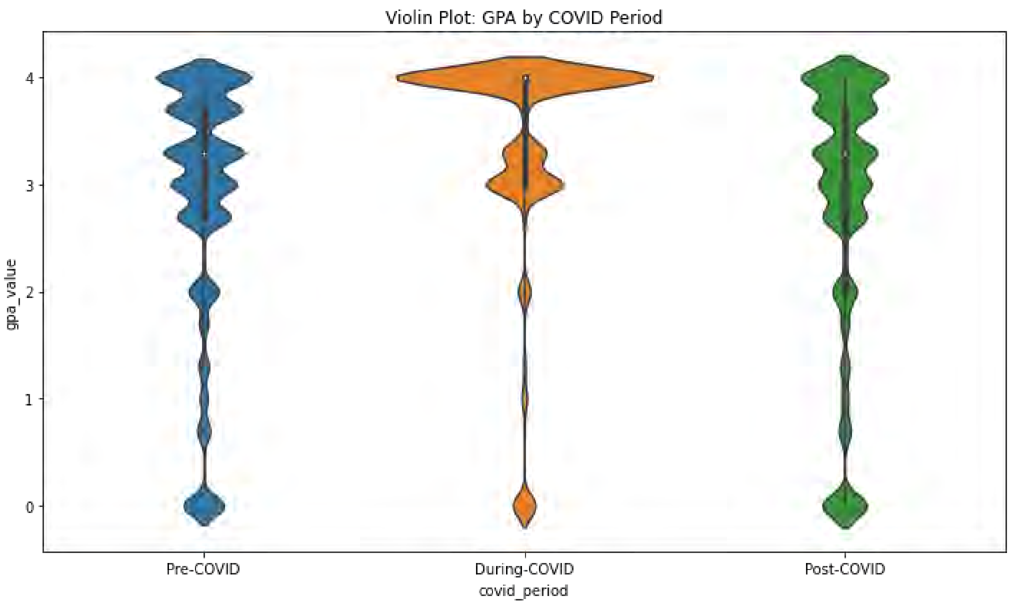

Figure 4.34: Violin plot of GPA distributions by COVID period. GPA clusters near the higher end During-COVID.

4.5.2 Statistical Model Results

Table 4.30 presents the results of the one-way ANOVA, which tests for significant differences in GPA across the three COVID periods. The ANOVA yields a highly significant **F-value of 12,250.58** ($p < 0.001$), confirming that GPA varies significantly across pandemic phases.

Table 4.30: One-Way ANOVA results for GPA across COVID periods. Significant differences detected ($p < 0.001$).

Source	df	Sum of Squares	F-value
COVID Period	2	34,700.18	12,250.58
Residual	825,491	1,169,116.00	–

Table 4.31 summarizes the results of the Tukey HSD post-hoc test, which identifies where specific GPA differences lie between periods. The largest mean difference of **-0.62** is observed between During-COVID and Post-COVID, followed by a difference of **-0.56** between During-COVID and Pre-COVID. Both differences are statistically significant ($p < 0.001$), highlighting the temporary GPA elevation during the pandemic. A smaller but still significant difference of **0.07** is found between Post-COVID and Pre-COVID.

Table 4.31: Tukey HSD post-hoc test results for GPA differences between COVID periods.

Group 1	Group 2	Mean Diff.	p-value	Lower	Upper
During–COVID	Post–COVID	-0.6228	<0.001	-0.6325	-0.6130
During–COVID	Pre–COVID	-0.5569	<0.001	-0.5661	-0.5477
Post–COVID	Pre–COVID	0.0658	<0.001	0.0591	0.0726

Table 4.32 presents the results of the multiple linear regression model, which tests the effects of COVID periods on GPA. The intercept value of **2.85** represents the baseline GPA in the Pre-COVID period. The coefficient for During-COVID is **0.5569** and statistically significant ($p < 0.001$), confirming a substantial increase in GPA during that period. Interestingly, Post-COVID shows a slight decline of **-0.0658** compared to Pre-COVID, again significant at the $p < 0.001$ level. These findings confirm a sharp GPA elevation during the pandemic, followed by a partial return to previous standards.

Table 4.32: Multiple Linear Regression results for GPA across COVID periods.

Predictor	Coef.	Std. Err	t-value	p-value
Intercept (Pre–COVID)	2.8533	0.002	1573.66	<0.001
During–COVID	0.5569	0.004	141.69	<0.001
Post–COVID	-0.0658	0.003	-22.75	<0.001

Table 4.33 displays the results from the mixed-effects model, which accounts for repeated observations within students by including random intercepts. The coeffi-

cient for During-COVID is **0.639**, the highest among all models, with a p-value $<$ 0.001, indicating a strong and significant increase in GPA. The Post-COVID period shows a smaller positive effect (**0.047**), still statistically significant. The group-level variance is estimated at **0.808**, reflecting meaningful individual-level differences in GPA trajectories.

Table 4.33: Mixed effects model results for GPA across COVID periods.

Predictor	**Coef.**	**Std. Err**	**z-value**	***p*-value**
Intercept (Pre–COVID)	2.672	0.005	495.60	<0.001
During–COVID	0.639	0.004	166.11	<0.001
Post–COVID	0.047	0.004	10.83	<0.001
Group Var (student) = 0.808				

Table 4.34 shows the Generalized Estimating Equations (GEE) results, which handle intra-student correlation using an exchangeable covariance structure. The During-COVID effect is consistent with previous models, estimated at **0.6365** ($p < 0.001$), while the Post-COVID effect remains modest at **0.0437**, also significant. This confirms the temporary GPA elevation during the pandemic and post-pandemic normalization.
Mixed-effects and GEE models, which account for within-student correlation, yield a small positive Post-COVID effect, whereas the pooled OLS model shows a small negative Post-COVID difference. Magnitudes are minimal in all cases, indicating sensitivity to model structure.

Table 4.34: GEE regression results with exchangeable covariance structure for GPA.

Predictor	**Coef.**	**Std. Err**	**z-value**	***p*-value**
Intercept (Pre–COVID)	2.6814	0.006	427.77	<0.001
During–COVID	0.6365	0.007	91.37	<0.001
Post–COVID	0.0437	0.009	5.11	<0.001

Table 4.35 displays the Kruskal-Wallis test result, a non-parametric test that does not assume normality. The high **H-statistic of 45711.71** and the p-value $<$ 0.001 confirm that GPA distributions differ significantly across the three COVID periods, corroborating previous findings.

Table 4.35: Kruskal-Wallis test result for GPA across COVID periods.

Statistic	**Value**
H-statistic	45711.71
p-value	<0.001

Table 4.36 summarizes the chi-square test on grade distributions across the three COVID phases. With a chi-square statistic of **111,577.74**, degrees of freedom 22, and $p < 0.001$, the test confirms a significant change in grade patterns. This aligns

with visual evidence showing increased frequencies of higher letter grades, especially A and A+, during the pandemic.

Table 4.36: Chi-square test results for grade distribution across COVID periods.

Statistic	**Value**
Chi-square Statistic	111,577.74
Degrees of Freedom	22
p-value	<0.001

4.5.3 Summary of Findings for Hypothesis 5

The results strongly support Hypothesis 3, confirming that student GPA and grade distributions varied significantly across the three COVID phases. Key findings include:

- **During-COVID GPA Surge:** GPA reached its highest levels during the pandemic, likely due to lenient grading policies, remote learning accommodations, and modified assessment strategies.
- **Post-COVID Reversion:** A slight GPA decline post-pandemic suggests reintroduction of stricter evaluation standards and normalization of academic rigor.
- **Consistent Statistical Evidence:** All models-ANOVA, OLS, mixed-effects, GEE, and Kruskal-Wallis-confirmed the statistical significance of GPA shifts across periods.
- **Grade Distribution Shift:** The chi-square test and heatmaps revealed a substantial increase in high grades (A, A+) during COVID, reinforcing the possibility of temporary grade inflation.

These findings indicate that the COVID-19 pandemic had a clear temporal impact on academic performance, with implications for future assessment and policy frameworks.

4.6 Hypothesis 6: Class Size and GPA

4.6.1 Hypothesis and Research Objective

Hypothesis 6: Class size impacts student grades, with smaller classes fostering better outcomes due to increased engagement and personalized attention.
Research Question: How do variations in class size influence academic outcomes in different disciplines?

4.6.2 Overview of Class Size and GPA Relationship

To examine the influence of class size on academic performance, we conducted correlation and linear regression analyses across the entire dataset, across programs, and at the individual course level.
The scatter plot in Figure 4.35 visualizes the relationship between class size and GPA. Despite a wide range of class sizes and considerable variation in GPA, a slight upward trend is observed. The linear regression line is superimposed in red to highlight this trend.

Figure 4.35: Scatter Plot: Class Size vs GPA (with Linear Fit)

At the aggregate level, the Pearson correlation between class size and GPA is $r = 0.0759$ with a p-value of < 0.0001, indicating a statistically significant yet weak positive relationship. The regression model yields the equation:

$$\text{GPA} = 0.0045 \times \text{Class Size} + 2.7591, \quad R^2 = 0.0058$$

This suggests that while the association is statistically significant, class size alone accounts for less than 1% of the variance in GPA.

4.6.3 GPA Distribution by Class Size Bands

To further understand this pattern, we grouped class sizes into bands of 5 and visualized GPA distributions within these groups using a boxplot (Figure 4.36). As shown, GPA appears more consistent and slightly higher in medium-sized to larger class bands (e.g., 41–60 students), whereas smaller classes exhibit higher variability.

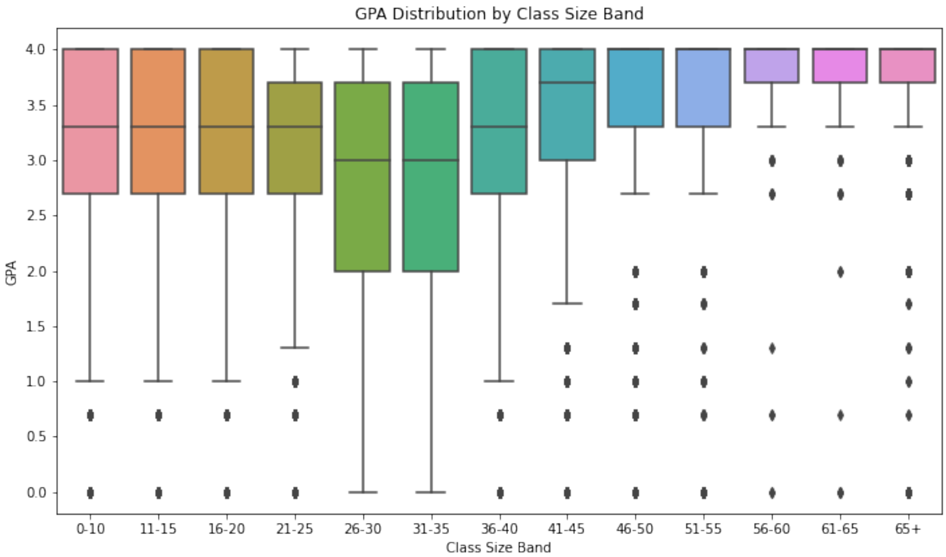

Figure 4.36: GPA Distribution Across Class Size Bands

4.6.4 Program-Level Analysis

We conducted separate analyses for major academic programs. Table 4.37 summarizes the correlation, regression slope, and explanatory power (R^2) across disciplines.

Table 4.37: Class Size vs GPA by Program

Program	Sample Size	Pearson r	Slope	R^2
MBA	1,960	0.1697	0.0184	0.0288
PHR	59,911	-0.1242	-0.0100	0.0154
BBA	156,792	0.1150	0.0084	0.0149
CS	66,391	0.1170	0.0046	0.0137
CSE	243,895	0.1173	0.0032	0.0127
ECO	31,407	0.0843	0.0033	0.0076
LLB	39,557	0.0844	0.0036	0.0071
APE	964	-0.0836	-0.0082	0.0070
ENG	31,643	0.0461	0.0058	0.0021
BIO	27,583	-0.0013	-0.0001	0.0000

The MBA program exhibited the highest $r = 0.1697$ and $R^2 = 0.0288$, suggesting a more pronounced class size effect. Conversely, BIO and ECE programs exhibited virtually no correlation.

4.6.5 Top Course-Level Correlations

Course-level analysis revealed several individual courses with notably strong correlations between class size and GPA. Table 4.38 lists the top 20 course-level results sorted by descending R^2.

Table 4.38: Top 10 Course-Level Correlations Between Class Size and GPA

Course Code	Sample Size	Pearson r	Slope	R^2
SOC420	104	0.6182	0.0201	0.3822
ANT415	84	0.4721	0.0196	0.2229
EEE423	383	0.4342	0.0491	0.1885
MKT424	1044	0.4300	0.0588	0.1849
SOC330	186	0.4051	0.0231	0.1641
CSE340	6497	0.3974	0.0742	0.1590
MKT428	170	0.3940	0.0458	0.1553
BTE311	100	-0.3697	-0.0438	0.1367
CSE250	6506	0.3512	0.0556	0.1233
EEE371	100	-0.3385	-0.0216	0.1146

Courses like `SOC420`, `ANT415`, and `CSE340` demonstrate a strong positive correlation, suggesting that in these cases, higher class size may even be associated with higher GPA-an unexpected result that may stem from grading structures, course difficulty levels, or self-selection of high-performing students into popular electives. This can also imply that the instructor resorted to making easier assessments in order to save time and difficulty of handling such a large section.

4.6.6 Summary of Findings for Hypothesis 6

Overall, the analysis indicates a weak but statistically significant positive correlation between class size and GPA in the general case. The low R^2 values suggest that class size alone is not a strong predictor of academic performance, although the effect is more pronounced in some programs (e.g., MBA) and courses (e.g., SOC420, CSE340).
These findings challenge the conventional belief that smaller classes always yield better academic outcomes. Instead, the relationship appears to be context-dependent, shaped by course structure, discipline-specific pedagogy, and possibly grading norms. Future work could explore non-linear models, control for instructor effects, or include qualitative measures of engagement to gain deeper insights into the class size-performance dynamic.

4.7 Hypothesis 7: Influence of Campus (RS vs Main) on GPA

Hypothesis: Student usage of the Residential Semester (RS) correlates with academic performance, with optimal usage times contributing to better outcomes.
Research Question: How does RS usage affect student grades, and are there differences based on department or program?

To explore whether campus location-Main Campus vs. RS (Residential Semester)-affects academic performance, we compared GPA distributions using descriptive and inferential statistics, visualizing the data using bar plots, box plots, and distribution curves.

Figure 4.37 shows the mean GPA of students in each campus type along with their standard deviations. RS students had a significantly higher mean GPA ($\mu = 3.286$, $\sigma = 0.689$) than Main Campus students ($\mu = 2.908$, $\sigma = 1.209$), based on $n = 68503$ and $n = 822,495$ respectively.

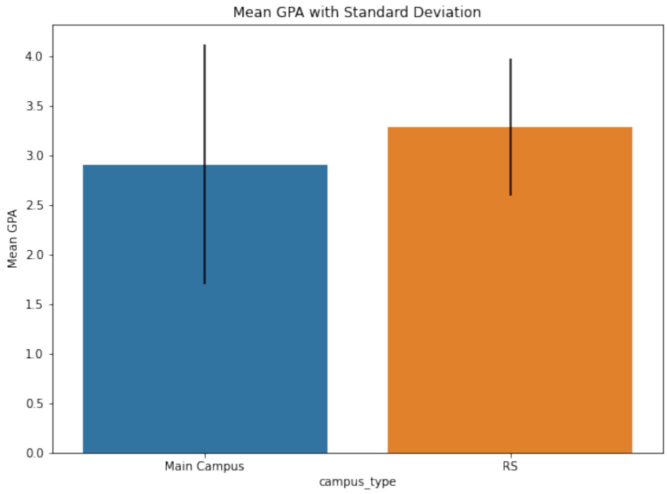

Figure 4.37: Mean GPA with Standard Deviation by Campus Type

Note: These n values are course-enrollment records, not unique students; per-student repeated-measures models (GEE/LMM) are reported separately. **Figure 4.38** presents the GPA distribution by campus using box plots. RS students exhibited higher median GPA and narrower interquartile range (IQR), suggesting better and more consistent performance.

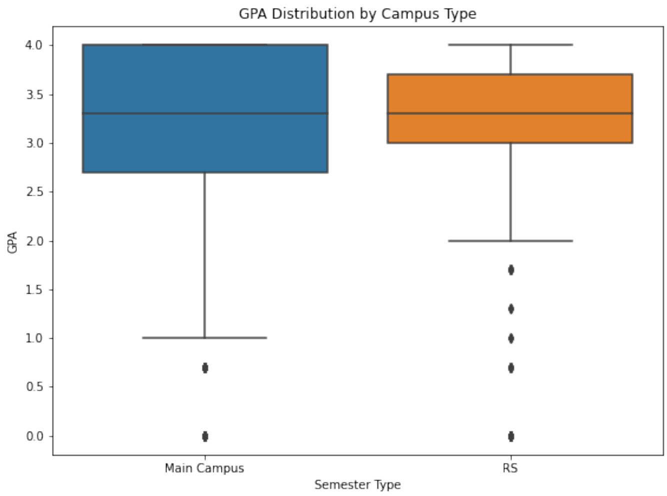

Figure 4.38: GPA Distribution by Campus Type (Box Plot)

Figure 4.39 shows the full GPA distribution via histograms and KDE curves. The RS GPA distribution is right-skewed and more concentrated around higher GPAs.

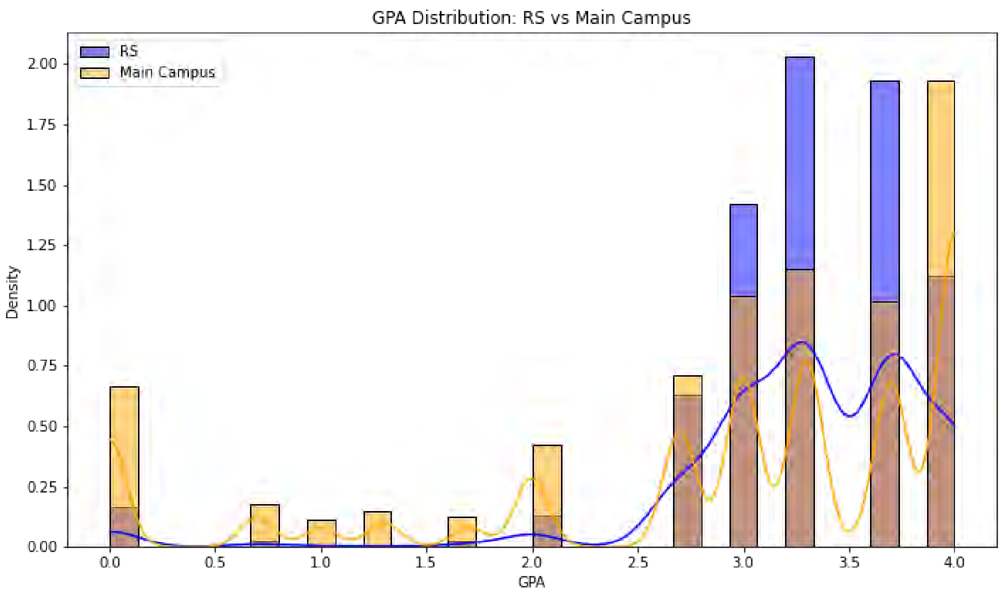

Figure 4.39: GPA Distribution Curves: RS vs. Main Campus

Table 4.39 provides the results of Welch's t-test and Mann-Whitney U test, confirming that RS students significantly outperform their Main Campus peers (all $p < .001$).

Table 4.39: Statistical Test Results: RS vs. Main Campus GPA

Test	**Statistic**	***p*-value**
Welch's T-Test	29.8674	2.87×10^{-172}
Mann-Whitney U	1,352,875,789.5	1.32×10^{-20}

Table 4.40 compares average GPAs by program for both campus types. Except for a few small programs, RS students consistently had higher GPA than Main Campus students.

Table 4.40: Average GPA by Program and Campus Type

Program	Main Campus GPA	RS GPA
ANT	3.023	2.268
APE	2.760	3.900
ARC	2.941	3.395
BBA	2.860	3.264
BIO	3.304	3.551
CS	2.856	3.257
CSE	2.881	3.278
ECE	2.489	3.539
ECO	3.078	3.533
EEE	2.678	3.342
EMBA	3.567	–
ENG	2.973	3.253
LLB	2.777	3.130
MAT	3.071	3.650
MBA	3.161	–
MIC	3.146	3.123
MS in BIO	2.567	–
MSCSSE	2.000	–
PHR	3.202	3.273
PHY	3.076	3.300

4.7.1 Summary of Findings for Hypothesis 7

The results clearly support Hypothesis 7. Students using RS campus facilities tend to have higher academic performance compared to their Main Campus counterparts. The significance is consistent across statistical tests and evident in multiple visualization techniques. While causality cannot be claimed, the findings highlight the potential positive role of RS environments in fostering better academic outcomes. Further research can explore the behavioral, logistical, and environmental aspects influencing this performance gap.

4.8 Hypothesis 8: Failure Trends and At-Risk Courses

Hypothesis Certain courses have consistently higher failure rates due to specific challenges (e.g., curriculum complexity, assessment design).
Research Question: What factors contribute to high failure rates in specific courses, and how can these be addressed?

To examine this hypothesis, we conducted a multi-level failure analysis by tracking trends in yearly failure rates, identifying course-level and program-level risks, and visualizing failure patterns over the past two decades. The findings reveal both systemic and course-specific challenges contributing to student underperformance.

Overall Yearly Failure Rate: Figure 4.40 illustrates the aggregate failure rate across all courses from 2005 to 2023. A steep decline from 18.1% in 2005 to around 8-10% between 2010 and 2018 is evident, suggesting gradual academic improvement. Notably, there is a dramatic dip in 2019-2020, possibly due to pandemic-induced grading flexibility, followed by a sharp rebound in 2022, indicating a return to stricter evaluation standards or post-pandemic academic disruption.

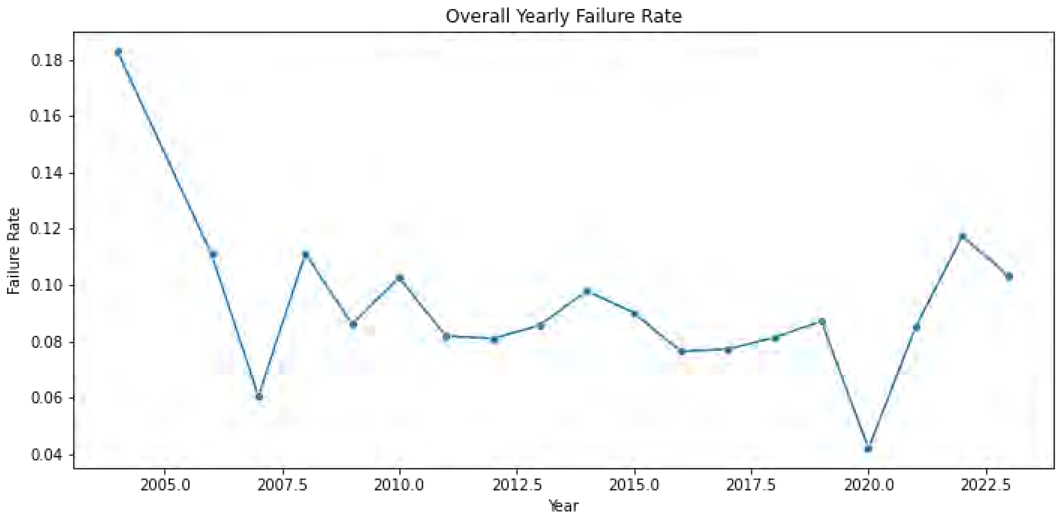

Figure 4.40: Overall yearly failure rate across all courses from 2005 to 2023. A significant decline is observed after 2005, with stabilization and a later surge in 2022 post-pandemic.

Failure Dispersion by Course: Figure 4.41 shows a scatter plot where each dot represents a course's failure rate in a given year. This visual highlights the persistent spread of outliers, some of which reach a 100% failure rate in individual years-indicating courses with extremely small enrollments or unique delivery issues.

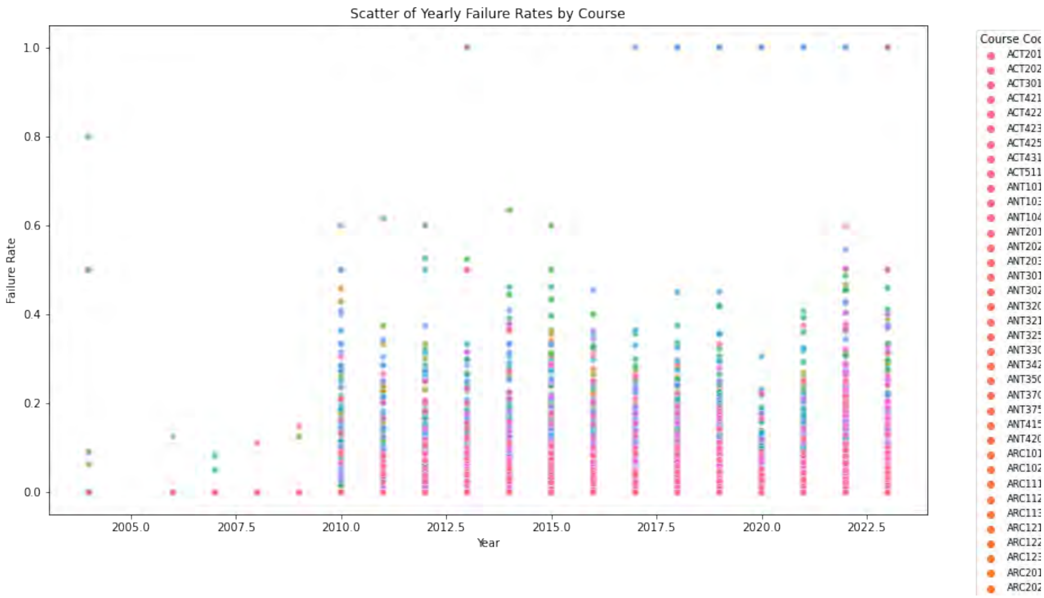

Figure 4.41: Scatter plot of yearly failure rates for individual courses. Persistent high-failure outliers suggest systemic difficulty or delivery challenges in specific courses.

Total Failures Over Time: Figure 4.42 depicts the total number of F grades issued per year. From a negligible count in early years, the total failures escalate significantly, peaking in 2022 with over 12,000 failing instances. This growth aligns with increased enrollment, evolving curricula, and possibly shifts in academic rigor.

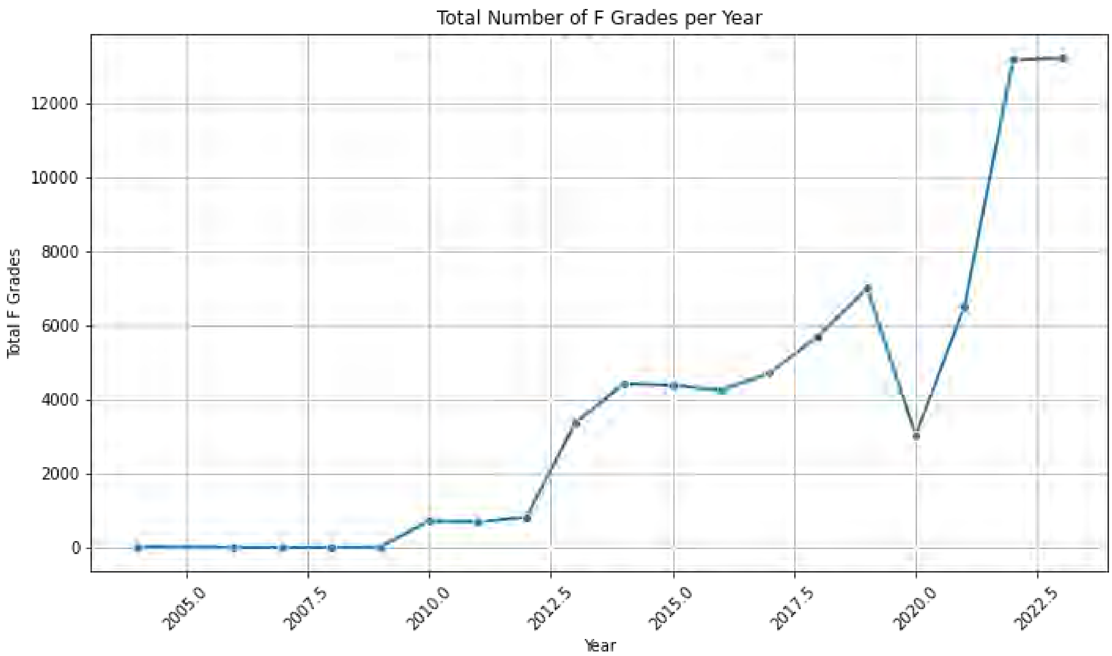

Figure 4.42: Total number of F grades issued each year. The data shows significant growth in failures, particularly in recent years with record-high counts post-2020.

Yearly Failure Rate by All Courses: Figure 4.43 tracks the yearly failure rate for all courses with at least 100 enrollments. The dense and overlapping trend lines indicate widespread fluctuation, yet a consistent cluster of high-failure courses stands out across years-warranting closer examination and targeted intervention.

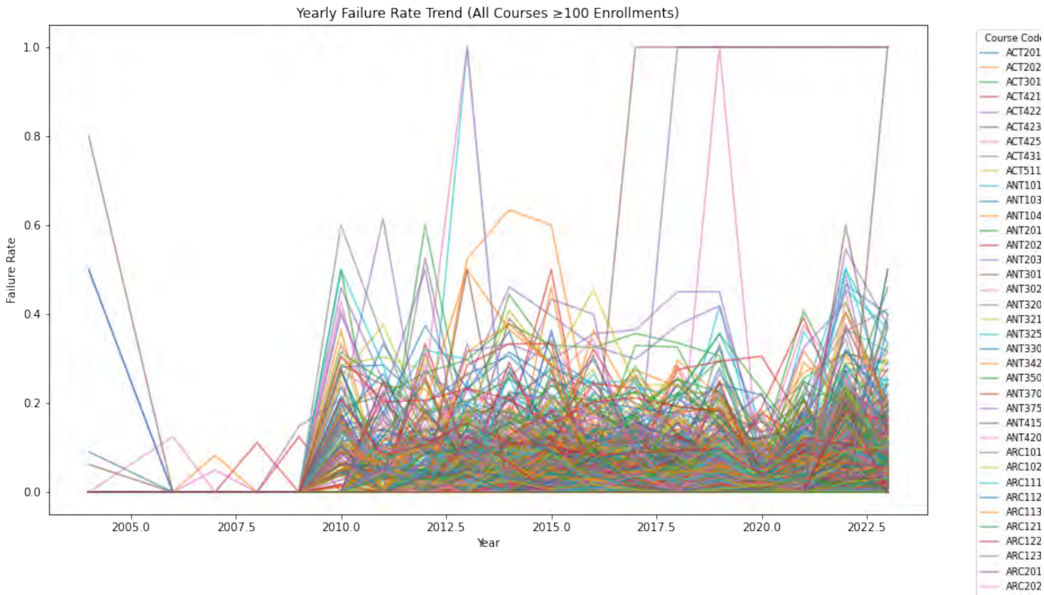

Figure 4.43: Failure rate trends for all courses with ≥100 enrollments. Several courses show consistently high failure trajectories over the years.

Top 5 High-Failure Courses: Figure 4.47 focuses on five consistently high-failure courses: CSE220, MAT110, CSE110, CSE111, and ENG091. These foundational and core courses show annual failure rates between 10-40%, confirming long-standing academic challenges. CSE220 in particular displays persistent peaks, reflecting its difficulty or possibly gaps in prerequisite preparation.

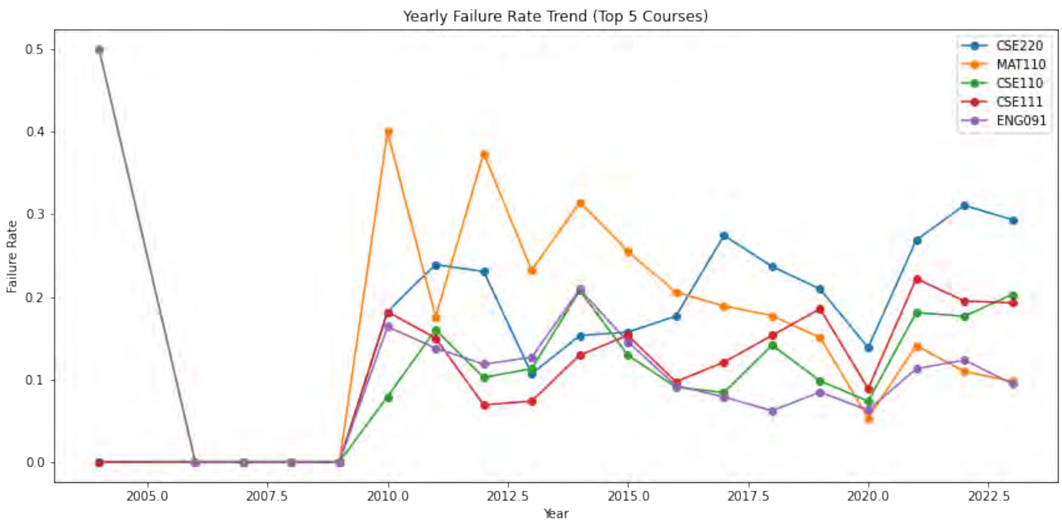

Figure 4.44: Yearly failure rate trends for top 5 high-failure courses. The consistent peaks validate these courses as systemic academic bottlenecks.

Program-Wise Failure Heatmap: Figure 4.45 presents a heatmap showing how failure rates of the top 15 high-risk courses vary across student programs. Some courses (e.g., CSE220, MAT110) affect multiple programs uniformly, while others (e.g., PHY111, CSE421) are program-specific. The heatmap allows departments to pinpoint which course-program combinations need curricular revision or additional support services.

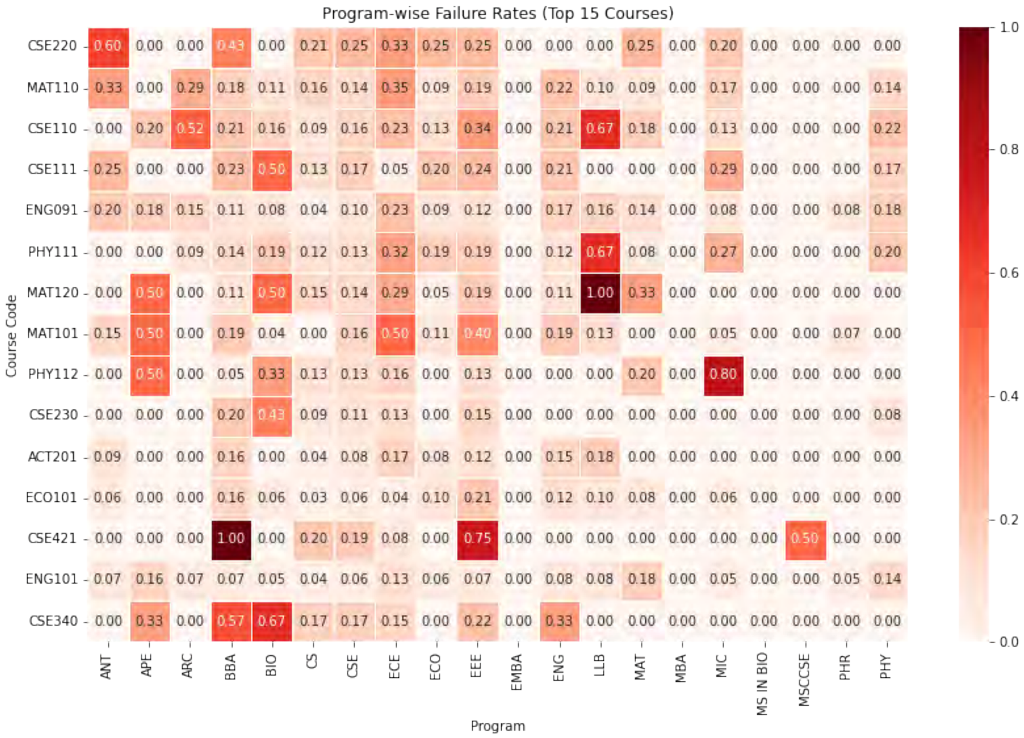

Figure 4.45: Heatmap showing program-wise failure rates for the top 15 high-failure courses. Shades of red indicate the severity of failure per program.

Top 20 Courses by Failure Rate: Table 4.41 summarizes the 20 courses with the highest overall failure rates. Courses like MAT092 and MAT091 show extremely high rates (97.7% and 75.7% respectively), possibly due to student placement errors or inadequate preparatory instruction. Other STEM-heavy courses (e.g., EEE201, CSE220, PHR104) indicate sector-wide academic stress areas. The counts reinforce the urgency for academic monitoring and instructional reform.

Table 4.41: Top 20 Courses by Failure Rate

Course Code	Fail Count	Total Count	Failure Rate
MAT092	897	918	0.977
MAT091	259	342	0.757
ECE210	47	111	0.423
EEE205	172	458	0.376
PHR103	68	209	0.325
LAW201	454	1427	0.318
EEE201	737	2383	0.309
EEE241	675	2228	0.303
ECE220	27	103	0.262
EEE101	146	564	0.259
PHR104	55	217	0.253
CSE220	2841	11610	0.245
ECE310	25	110	0.227
MAT104	200	916	0.218
ECO104	145	682	0.213
EEE301	316	1588	0.199
PHR204	68	342	0.199
PHY113	21	106	0.198
CSE421	1134	5846	0.194
EEE209	311	1606	0.194

4.8.1 Summary of Findings for Hypothesis 8

Our analysis highlights a persistent subset of courses where students face significant academic difficulty. These findings point to the need for curriculum redesign, prerequisite enforcement, enhanced instructional quality, and targeted support interventions in consistently high-failure courses. Identifying and supporting these at-risk courses is essential for improving student progression and overall academic health.

4.9 Hypothesis 9: Program- and Department-Level Performance Differences

Research Question (RQ9): Are there programs or departments where students consistently perform better or worse, and what factors contribute to these patterns?
Hypothesis (H9): Statistically significant performance variation exists across programs and departments, reflecting the combined influence of delivery mode, curriculum structure, assessment design, and environmental factors identified across H1-H8.

Purpose and Rationale. Building upon the preceding nine hypotheses, this hypothesis moves beyond isolated factor-level analyses to assess how those influences interact at the departmental level. While H1-H8 examined delivery mode, class size, prerequisite alignment, RS environment, and other structural dynamics individually, H9 investigates how these effects *accumulate and differ* across programs. It thus captures the university's academic ecosystem as a system of interconnected yet diverse units-each responding differently to structural, technological, and pedagogical factors.

In doing so, H9 bridges the gap between statistical outcomes and institutional meaning. It transforms the previously identified factor-level relationships into a broader diagnostic lens-one that can guide university-wide academic benchmarking, policy decisions, and departmental improvement strategies.

4.9.1 Summary of Findings and Interpretation

ANOVA and Kruskal-Wallis tests confirmed statistically significant GPA variation across both programs and departments ($p < 0.001$). Post-hoc analyses indicated a stable rank-order of departmental GPA means, with patterns remaining consistent over multiple years. Departments with more structured curricula, smaller classes, and coherent prerequisite scaffolding (e.g., ENH, SoL) exhibited higher and more consistent GPA outcomes.

Conversely, resource-intensive and STEM-focused units (e.g., CSE, EEE, Physical Sciences) recorded lower means and greater variance-reflecting stricter grading norms, heavier prerequisite dependencies, and higher failure concentration in gateway courses.

The results show that departmental outcomes are not random but arise from the unique combination of factors previously modeled in H1-H8. For example, pandemic-related delivery shifts (H1, H4) benefited humanities and law programs, whereas large-section technical courses (H5) experienced performance dips. Similarly, strong prerequisite dependencies (H3) and recurring gateway bottlenecks (H8) intensified grade dispersion in STEM courses, while RS exposure (H6) enhanced stability for certain programs.

These cross-hypothesis interactions reveal that institutional performance variation is fundamentally *discipline-specific*-each department embodies a different configuration of the same structural influences.

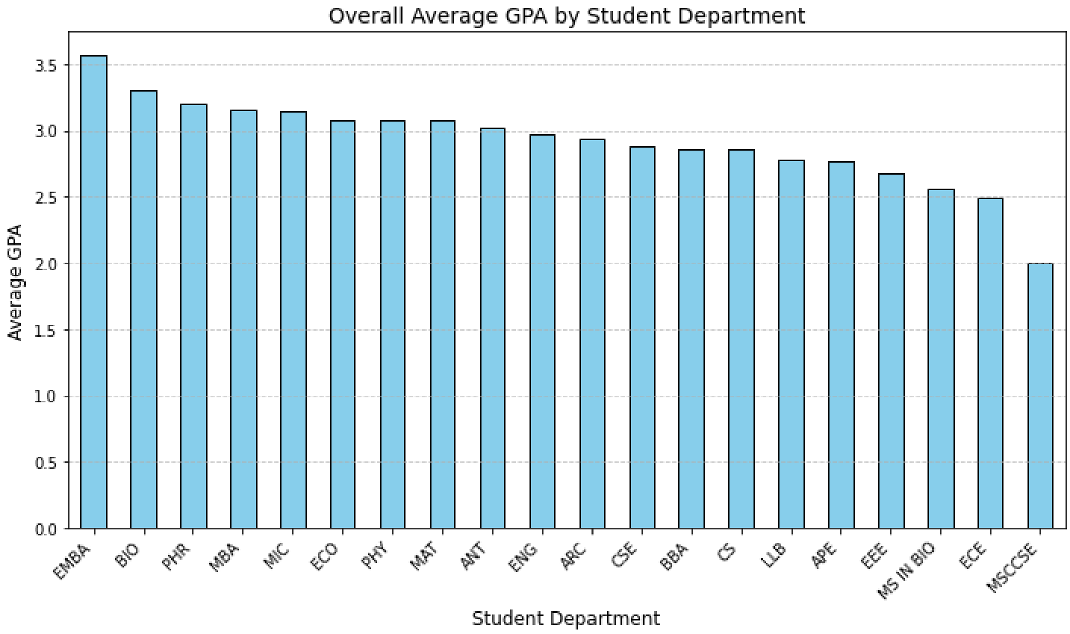

Figure 4.46: Program-wise Average GPA

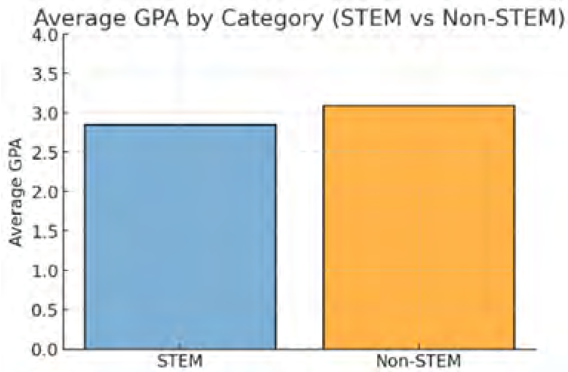

Figure 4.47: STEM vs Non-STEM Course Average GPA

Table 4.42: Differential Departmental Patterns and Structural Drivers

Department / Program	**Observed Trend**	**Likely Structural Drivers (Link to H1–H8)**
English and Humanities (ENH)	High mean GPA; low grade dispersion	Benefited from hybrid/online flexibility (H1, H4); smaller class sizes (H5); coursework-based continuous assessment
School of Law (SoL)	Above-average GPA; consistent across COVID phases	Structured curriculum and strong RS exposure (H6); limited gateway dependencies (H2, H3)
Business / Economics	Mid-range GPA; moderate variance	Balanced delivery mix (H1); manageable class size (H5); fewer high-failure hotspots (H7)
Computer Science and Engineering (CSE)	Lower GPA; higher grade dispersion	Large class sizes (H5); strong prerequisite dependencies (H3); recurring gateway failures (H7); pandemic assessment transition (H4)
Electrical and Electronic Engineering (EEE)	Lower GPA; high variance across terms	Rigor in assessment design (H2); lab-intensive courses; cumulative prerequisites (H3); high section sizes (H5)
Physical Sciences (MAT, PHY)	Lower mean GPA; steep declines in core terms	Heavy prerequisite dependency (H3); limited hybrid adaptation (H1); concentrated failure hotspots (H7)
RS-linked Programs	Higher and more consistent GPA	Structured mentoring and peer learning (H6); immersive environment reduces performance variability

Interpretation. Performance disparities across departments arise from discipline-specific combinations of structural, pedagogical, and contextual conditions. Humanities-oriented programs benefited from adaptive delivery and assessment flexibility, while technical and quantitative disciplines experienced greater volatility due to intensive prerequisite structures and large cohort sizes. These findings confirm that the institutional landscape of academic performance is patterned-not by chance, but by the differential expression of the same underlying factors examined throughout H1-H8.
Conclusion. The combined evidence validates that inter-departmental GPA variation is real, measurable, and structurally driven. Rather than treating each factor in isolation, this synthesis illustrates how delivery environment, curriculum architecture, and course dependencies collectively shape departmental outcomes. These results support the establishment of a *Departmental Performance Benchmarking Framework*-a continuous monitoring tool that identifies strong pedagogical models, isolates systemic weaknesses, and advances academic equity across disciplines.

4.10 Discussion

This study investigated nine interrelated hypotheses to understand the structural, pedagogical, and contextual determinants of academic performance at BRAC University. Taken together, these hypotheses form an integrated view of how delivery environment, curriculum design, class size, and learning context collectively shape student outcomes. Each hypothesis contributes a specific layer of insight-from course-level predictors to system-level disparities-culminating in an institutional understanding of performance variation across programs and departments.

Hypothesis 1 examined the effect of delivery mode on student performance. The analysis revealed that online and hybrid delivery modes were associated with higher mean GPA and reduced variance compared to traditional offline courses. This suggests that well-structured digital learning environments may enhance student outcomes through flexible pacing, clearer assessment structures, and increased accessibility. However, mode effects also varied across disciplines-humanities and law programs benefited more strongly from online adaptation, whereas STEM departments displayed more conservative gains due to laboratory and design-based learning constraints.

Hypothesis 2 explored course-level influence on overall academic success. Using a combined metric of correlation and enrollment frequency, the study identified high-impact courses such as CSE110, MAT110, and ENG101 as critical determinants of cumulative GPA. These results confirm that a small set of foundational courses disproportionately shape long-term performance trajectories, emphasizing the value of targeted curricular investment in gateway subjects.

Hypothesis 3 focused on the relationship between prerequisite and core courses, finding strong and statistically significant positive correlations across most course pairs ($r \approx 0.4 - 0.6$, $p < 0.001$). This confirms that mastery of prerequisite content is a powerful predictor of subsequent success, particularly within STEM disciplines. Departments with well-aligned prerequisite-core sequences exhibited more stable grade distributions, while weaker alignments led to higher grade dispersion.

Hypothesis 4 investigated the institutional effects of the COVID-19 pandemic, uncovering substantial temporal and departmental variation. During the pandemic, GPAs increased modestly across several departments-especially ENH and SoL-indicating adaptive grading practices or flexible assessments. In contrast, technical disciplines such as CSE and EEE experienced declines in post-pandemic terms, reflecting challenges in transitioning practical and design-intensive coursework online. These differences highlight that the pandemic's educational effects were not universal but mediated by the pedagogical nature of each field.

Hypothesis 5 examined temporal performance shifts across pandemic phases. The data revealed clear adaptation and recovery patterns: performance peaked during fully online semesters and normalized in post-COVID periods as traditional assessments resumed. This trend reinforces that environmental disruptions can trigger systemic academic responses, with delivery flexibility temporarily mitigating performance barriers.

Hypothesis 6 analyzed class size effects on academic outcomes. Although the overall correlation between class size and GPA was weak, department-level results revealed context-specific nuances. In humanities and social science programs, smaller class sizes correlated with slightly higher GPA and reduced grade variance, while in high-enrollment technical courses, the relationship was negative-larger cohorts often corresponded to reduced consistency and elevated failure rates. These findings suggest that the class size effect is conditional, moderated by teaching style, subject complexity, and instructional load.

Hypothesis 7 compared academic performance between Residential Semester (RS) and Main Campus contexts. RS students achieved significantly higher and more consistent GPAs ($\mu_{RS} = 3.286$, $\mu_{Main} = 2.908$, $p < 0.001$). The structured, immersive environment of the RS-with continuous faculty engagement and peer learning-appears to cultivate academic discipline and performance consistency. Replicating similar mentoring and community-learning models in the Main Campus could yield institution-wide benefits.

Hypothesis 8 addressed persistent high-failure courses, identifying repeated bottlenecks in gateway and foundational subjects such as CSE220, EEE201, and MAT092. Long-term trends showed failure-rate spikes during the pandemic followed by a gradual normalization, yet the same set of courses consistently re-emerged as performance hotspots. This pattern points to systemic challenges in course design and prerequisite scaffolding rather than temporary instructional lapses. Early-warning analytics and targeted instructional support for these courses are recommended to reduce academic attrition.

Finally, **Hypothesis 9** synthesized the preceding findings to explore how these structural effects vary across programs and departments. The results demonstrated statistically significant GPA differences between departments (ANOVA and Kruskal -Wallis, $p < 0.001$), revealing that each discipline expresses institutional factors in unique ways. Humanities-oriented programs (e.g., ENH, SoL) achieved higher and more stable GPAs, benefiting from hybrid adaptability, smaller cohorts, and continuous evaluation. In contrast, STEM and quantitative departments (e.g., CSE, EEE, Physical Sciences) recorded lower averages and higher variance, reflecting rigorous assessment norms, heavier prerequisite dependencies, and recurring high-failure gateways. Business and Economics programs maintained mid-range stability, balancing course load and delivery diversity, while RS-linked programs exhibited consistently higher performance due to structured mentoring and engagement.

These patterns confirm that performance variation across academic units is not random but structurally patterned. Each department embodies a distinct configuration of the same underlying influences identified in H1-H8-delivery mode, class size, prerequisite design, and learning environment-manifesting differently depending on disciplinary context. The findings thus provide an institutional "performance map" that can inform strategic resource allocation, benchmarking, and pedagogical reform.

Overall Discussion. Collectively, the nine hypotheses converge on a central conclusion: student achievement in higher education emerges from a complex interplay between institutional design, instructional practice, and contextual environment. Academic performance is not solely a reflection of individual effort but of how the university's structural mechanisms-curriculum coherence, teaching models, delivery modes, and assessment frameworks-interact across time and disciplines. Future institutional strategies should therefore adopt a data-driven and system-aware approach: (1) strengthen high-impact and prerequisite courses, (2) replicate RS-style engagement in non-residential settings, (3) reduce large-class disparities, and (4) establish departmental benchmarking frameworks to monitor performance equity and quality across programs.

This holistic perspective transforms isolated findings into a cohesive, evidence-based understanding of how educational structure shapes academic success within a complex institutional ecosystem. The impact ranking across hypotheses reinforces this systemic view. The strongest influences (H3 and H2) lie in the curriculum's internal logic—how prerequisite mastery and key gateway courses determine long-term academic outcomes. The next layer (H8 and H7) reflects institutional equity and structural pressure points, such as high-failure zones and environmental factors. Contextual variables (H4, H1, and H6) serve as amplifiers—important but conditional on pedagogical design—while operational factors like class size (H5) exhibit the weakest standalone effect. This hierarchy of influence suggests that curricular integrity and instructional scaffolding drive sustainable academic performance, whereas contextual and operational refinements enhance but cannot substitute for strong academic foundations.

4.11 Recommendations

Across all nine hypotheses, the findings of this study highlight the transformative potential of **data-informed policymaking** in higher education. The results demonstrate that student performance is not solely a product of individual aptitude but is also shaped by institutional design, pedagogical structure, and contextual learning environments. Translating these insights into practice requires a systematic framework through which empirical evidence continually informs academic decision-making.

By embedding analytics-driven evaluation into curriculum development, instructional design, and resource planning, universities can evolve from reactive to adaptive institutions-ones capable of anticipating challenges and optimizing learning outcomes proactively. The proposed recommendations collectively advocate for a model of **continuous academic optimization**, where data insights guide course improvement, teaching quality enhancement, and equitable policy implementation.

To sum up, this study underscores the importance of maintaining a feedback loop between research and institutional practice. When educational policy becomes responsive to real performance data and contextual evidence, higher education can achieve both academic excellence and structural resilience. Ensuring that every program, course, and learner benefits from an evidence-based approach to educational advancement.

4.11.1 H1 - Impact of Learning Environment (Online, Hybrid, and In-Person)

The comparison among online, hybrid, and offline learning environments revealed distinct performance patterns, with most programs showing progressive GPA improvement from in-person to hybrid to fully online settings. However, these trends were not uniform across disciplines, as certain programs-particularly advanced or laboratory-based courses-experienced more modest gains or even slight declines. This divergence suggests that learning mode interacts deeply with course type, pedagogy, and assessment design.

It should also be acknowledged that grading policies during the fully online semesters were temporarily relaxed, with thresholds for top grades being lowered (e.g., marks around 70 qualifying for an "A", which would typically correspond to a "B-" under regular conditions). Such contextual adjustments likely amplified the observed GPA elevation in online periods, underscoring the importance of interpreting these results within their institutional framework.

Based on these findings, the university should move beyond a binary conception of "online versus offline" and adopt a blended learning optimization framework that identifies which courses benefit most from sustained online or hybrid components. For example, theory-based or discussion-intensive subjects could retain asynchronous or hybrid elements, while laboratory and design courses remain predominantly in-person.

Faculty development programs should also be expanded to include specialized modules on digital pedagogy, focusing on student motivation, formative assessment in online contexts, and hybrid course design. Implementing a university-wide "Digital Pedagogy Fellowship" could further encourage innovation by recognizing instructors who effectively leverage technology to enhance learning outcomes. Finally, since student performance varied across modalities, continuous feedback mechanisms, such as post-semester surveys or analytics dashboards, should inform dynamic adjustments to delivery modes, ensuring pedagogical equity across programs.

4.11.2 H2 - Course-Level Contribution to CGPA

The strong correlation between performance in specific foundational courses and overall CGPA emphasizes their disproportionate influence on students' long-term academic trajectories. Courses such as introductory programming, mathematics, and communication skills act as academic "leverage points." The university should therefore prioritize these gateway courses in terms of teaching quality, curriculum review, and continuous improvement cycles. A High-Impact Course Task Force could be established to identify and monitor such courses annually, recommending targeted pedagogical interventions where needed.

Additionally, an Early Risk Detection System should be implemented through institutional analytics using real-time performance data to flag students underperforming in these critical subjects. Such systems could trigger automated academic advising, remedial sessions, or peer tutoring before cumulative CGPA impact becomes irreversible. Finally, the findings advocate for linking course-level evaluations not just to instructor performance but to long-term student progression metrics, thereby aligning pedagogical accountability with measurable academic outcomes.

This course-level monitoring strategy would transform individual course improvement into systemic academic advancement.

4.11.3 H3 - Prerequisite: Core Course Dependency

The prerequisite-core relationship analysis indicated that strong performance in foundational subjects significantly enhances success in subsequent core courses. This finding suggests that learning continuity is crucial, and disruptions at early stages have cascading effects. To strengthen this academic scaffolding, the university could conduct a curriculum coherence audit to ensure logical sequencing, workload balance, and alignment of learning outcomes across prerequisite chains.
Where weak prerequisite mastery correlates with subsequent failure, the university may introduce bridging modules or diagnostic mini-courses before students enroll in advanced subjects. These could be short, intensive refreshers delivered online during semester breaks. Furthermore, academic advisors should be equipped with data dashboards that visually track each student's performance trajectory across prerequisite paths, enabling more personalized academic guidance. Institutionalizing such preventive measures would ensure that progression through the curriculum remains both equitable and pedagogically sound.

4.11.4 H4 - Departmental Variation in COVID's Academic Impact

The analysis revealed that the pandemic's academic disruptions were not uniform across departments, underscoring disparities in adaptability, resource readiness, and pedagogical flexibility. This finding highlights the importance of departmental preparedness in the face of systemic shocks. To address such asymmetries, the university should consider institutionalizing a Departmental Digital Resilience Framework (DDRF) that evaluates each unit's capacity for remote or blended instruction. Departments that struggled with lower mean GPAs during the online transition could benefit from targeted capacity-building initiatives such as structured training on learning management systems (LMS), digital assessment integrity, and student engagement strategies in virtual settings.
Moreover, periodic "crisis-readiness audits" should be conducted to assess infrastructural and pedagogical adaptability, particularly for resource-dependent departments like architecture or laboratory-based sciences. These audits could inform a differentiated response plan in future emergencies, ensuring that academic continuity is sustained without widening inter-departmental inequities. By embedding resilience into departmental policy and resource allocation, the university would not only mitigate future disruptions but also strengthen long-term instructional flexibility across disciplines.

4.11.5 H5 - Temporal Effects of COVID-19 on Academic Recovery

The temporal analysis showed that student performance gradually stabilized in the semesters following the pandemic, but residual effects lingered for several cohorts. This suggests that while structural normalcy returned, underlying academic and

psychological disruptions persisted. The university could respond by embedding academic recovery protocols that provide adaptive support for post-crisis cohorts. These may include bridge modules to reinforce foundational knowledge gaps, mentorship programs pairing senior students with newer entrants, and flexible course-load options during recovery periods.
Moreover, the administration could collaborate with the Counseling Unit and Student Life Office to design integrated academic well-being frameworks ensuring that recovery is not confined to grades alone but extends to mental resilience and engagement. Departments may also use the data-driven insights from this study to plan staggered course scheduling or modified assessment loads in the semesters immediately following institutional disruptions. Such measures would foster a more sustainable return to academic stability and minimize long-term grade volatility across departments.

4.11.6 H6 - Class Size and Academic Outcomes

The inverse relationship observed between class size and academic performance reaffirms that instructional quality and student engagement are sensitive to cohort scale. This finding warrants a structural policy response. The university could establish evidence-based section size caps for both theory and lab courses, ideally maintaining smaller cohorts for courses emphasizing discussion, critical analysis, or complex problem-solving. Large-enrollment courses, on the other hand, should be complemented by parallel tutorial sessions or teaching assistant (TA) programs to preserve interactivity and feedback quality.
Administrative planning for semester scheduling could integrate predictive models that balance instructor availability, classroom capacity, and ideal section sizes derived from this study's data. Over time, departments should evaluate GPA distributions alongside section sizes as part of annual course performance reviews. By incorporating empirical evidence into capacity planning, the institution would align pedagogical design with measurable learning effectiveness rather than logistical convenience.

4.11.7 H7 - RS (Residential Semester) vs. Main Campus Performance

The comparison between RS and main-campus semesters revealed that the RS environment fosters stronger focus and consistency among students, particularly in early academic stages. This outcome suggests that the structural simplicity, peer closeness, and immersive nature of the RS model support better academic regulation. To replicate these benefits more broadly, the university might design micro-RS modules within the main campus framework - short-term, concentrated academic sessions with smaller groups and heightened faculty engagement.
Furthermore, RS semesters could serve as laboratories for pedagogical experimentation - testing alternative assessment formats or community-based learning models that could later scale to the main campus. The administration should also review resource allocation, ensuring equitable access to academic advising, mentoring, and co-curricular support for both RS and main-campus students. By leveraging the RS model as a controlled learning ecosystem, the university could derive replica-

ble strategies for enhancing academic discipline, social cohesion, and instructional quality across all programs.

4.11.8 H8 - Courses with High Failure Rates

Courses exhibiting persistently high failure rates represent critical nodes of academic vulnerability. Beyond indicating student difficulty, these patterns often signal issues in assessment design, content overload, or instructional mismatch. The university should institutionalize course-level performance audits every semester, analyzing grade distributions, instructor feedback, and student evaluations to identify underlying causes.

High-failure courses should be flagged for pedagogical redesign initiatives, incorporating modular content delivery, formative assessments, and learning support labs. Establishing Peer Learning Programs - where senior students mentor those repeating high-risk courses - could further reduce attrition. Departments should also be encouraged to pilot assessment diversification, moving away from heavy final-exam weighting toward continuous evaluation methods. Collectively, these measures would shift institutional focus from reactive remediation to proactive course improvement, enhancing both student retention and instructional accountability.

4.11.9 H9 - Program and Department Level Performance Differences

To address the performance disparities identified under H9 and foster greater academic equity across programs, the university should establish a *Departmental Performance Benchmarking Framework*. This framework would systematically monitor GPA trends, failure concentrations, and course-level leverage metrics each term, enabling data-driven comparisons across departments.

High-performing units can serve as institutional exemplars for instructional design, assessment calibration, and mentoring structures, while programs showing consistent underperformance should receive structured interventions-such as targeted teaching-development workshops, learning analytics dashboards, and enhanced prerequisite-bridging initiatives.

Regular cross-department reviews, anchored in standardized learning-outcome rubrics, can further harmonize grading practices, align curricular expectations, and promote consistent academic standards across the university.

4.11.10 Summary Remark

Across all hypotheses, the findings of this study highlight the potential for data-informed policymaking in higher education. By transforming empirical insights into structured interventions - from optimized class sizes and adaptive learning modes to targeted course redesign - the university can evolve toward a model of continuous academic optimization. This approach reinforces the role of institutional research not merely as a diagnostic exercise but as a dynamic mechanism for improving teaching, learning, and student success.

Chapter 5

Conclusion and Future Work

5.1 Conclusion and Future Work

This study set out to examine the complex and multifactorial influences shaping academic performance in higher education. By analyzing institutional data through nine carefully formulated hypotheses, it uncovered meaningful relationships between academic outcomes and a wide range of variables, including the COVID-19 pandemic, learning environments, prerequisite structures, course-level contributions, class sizes, campus resources, and failure patterns in specific courses.

The findings confirmed that the COVID-19 pandemic had a significant and uneven effect across disciplines, with some programs adapting more successfully than others to remote learning. While online delivery during the pandemic was associated with higher observed GPAs in aggregate—particularly in 2020–2021—effects varied by program; post-pandemic GPAs declined slightly yet remained above pre-pandemic levels. Foundational courses, particularly introductory programming and mathematics, were shown to exert substantial influence on cumulative CGPA, emphasizing the need for targeted support in these academic bottlenecks. The relationship between performance in prerequisite courses and subsequent core subjects further highlighted the importance of curriculum scaffolding. While earlier assumptions suggested that smaller classes would consistently yield higher academic outcomes, the empirical results indicate a more nuanced reality. The observed correlation between class size and GPA was weak and slightly positive ($r \approx +0.076$, $R^2 \approx 0.006$), suggesting that class size alone exerts minimal influence on performance. Differences across programs and course types appear to moderate this relationship—some programs, such as business-oriented tracks, displayed modest positive effects, whereas others showed negligible or even inverse trends. Overall, class size should therefore be interpreted as a contextual rather than deterministic factor in student achievement.

Beyond academic structures, the study revealed that students at the university's Residential Semester (RS) achieved consistently higher GPAs, underscoring the importance of well-being and holistic support in academic success. Failure-rate analyses identified a cluster of high-risk courses, particularly in STEM and technical disciplines, that require immediate pedagogical attention, enhanced instructional design, and additional student support mechanisms.

Taken together, these findings contribute to institutional decision-making by presenting a data-driven model that integrates academic, structural, and environmental perspectives. They suggest that universities should adopt multidimensional strategies that align curriculum design, student support services, and wellness initiatives in order to foster sustainable academic success.

Looking ahead, several opportunities for future research emerge. First, incorporating qualitative approaches such as student interviews and focus groups could provide richer insight into the behavioral, motivational, and emotional drivers behind the quantitative patterns observed. Second, expanding the dataset to cover additional academic years would enable longitudinal analyses, clarifying long-term effects of pandemic-era learning. There is also significant potential to design and evaluate targeted interventions for high-failure courses, including peer mentoring programs, flipped classroom approaches, and adaptive assessments, to assess their impact on resilience and retention. Further, validating this analytical framework across other institutions could yield comparative insights, highlighting both common challenges and context-specific differences. Finally, combining traditional statistical methods with interpretable machine learning could strengthen early detection of at-risk students and support more proactive advising and institutional policy planning.

Collectively, the nine tested hypotheses offered a cohesive picture of how delivery mode, curriculum structure, and institutional environment interact. The combined model demonstrated that academic performance is not determined by a single factor but by interlocking dynamics between pedagogy, structure, and context. This holistic framing provides a replicable framework for institutional analytics in comparable universities across South Asia.

In sum, this study provides a foundation for building data-informed, holistic strategies to improve academic performance, combining pedagogical innovation with structural and well-being considerations.

Bibliography

[1] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[2] B. L. Welch, "The generalization of student's problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1–2, pp. 28–35, 1947.

[3] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-way analysis of variance," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[4] K. J. Edwards, "Correcting partial, multiple and canonical correlations for the attenuation of measurement error," Johns Hopkins Univ. Center for the Study of Social Organization of Schools, Tech. Rep., 1971.

[5] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[6] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986. DOI: 10.1093/biomet/73.1.13.

[7] J. Cohen, "Statistical power analysis for the behavioral sciences," in *Classic Texts in the Behavioral Sciences*, Routledge, 1988.

[8] S. Siegel and N. J. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, 2nd. New York, NY: McGraw-Hill, 1988.

[9] L. G. Tornatzky and M. Fleischer, *The Processes of Technological Innovation.* Lexington Books, 1990.

[10] L. S. Aiken and S. G. West, *Multiple Regression: Testing and Interpreting Interactions*. Sage, 1991.

[11] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[12] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. New York, NY: John Wiley & Sons, 1999.

[13] W. H. DeLone and E. R. McLean, "The delone and mclean model of information systems success: A ten-year update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30, 2003.

[14] P. Auer and R. Meir, "Generalization bounds and statistical performance of cross-validation and fdr procedures," *Journal of Machine Learning Research*, vol. 5, pp. 397–422, 2004.

[15] G. Keppel and T. D. Wickens, *Design and Analysis: A Researcher's Handbook*, 4th. Upper Saddle River, NJ: Pearson Prentice Hall, 2004.

[16] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th. New York, NY: McGraw-Hill Irwin, 2005.

[17] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd. Hoboken, NJ: Wiley, 2007.

[18] G. D. Kuh, "What engagement data tell us about readiness," *Peer Review*, vol. 9, no. 1, pp. 4–8, 2007.

[19] M. M. Yusof, A. Papazafeiropoulou, R. J. Paul, and L. K. Stergioulas, "Investigating evaluation frameworks for health information systems," *International Journal of Medical Informatics*, vol. 77, no. 6, pp. 377–385, 2008.

[20] S. E. Carrell, R. L. Fullerton, and J. E. West, "Cohort and peer effects in college achievement," *Journal of Labor Economics*, vol. 27, no. 3, pp. 439–466, 2010.

[21] J. W. Osborne, "Improving your data transformations: Applying the box–cox transformation," *Practical Assessment, Research, and Evaluation*, vol. 15, no. 12, pp. 1–9, 2010, Includes discussion of disattenuation and reliability corrections.

[22] J. Monks and R. Schmidt, "The impact of class size on outcomes in higher education," *The BE Journal of Economic Analysis & Policy*, vol. 11, no. 1, 2011.

[23] K. N. et al., "The assumption of a reliable instrument and other pitfalls," *PMC*, 2012.

[24] J. W. Hardin and J. M. Hilbe, "Robust (sandwich) variance in gee," in *Generalized Estimating Equations*, Chapter on robust variance, Chapman & Hall/CRC, 2012.

[25] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC, 2012.

[26] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions* (Advanced Quantitative Techniques in the Social Sciences), 2nd. Thousand Oaks, CA: SAGE Publications, 2012.

[27] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 5th ed. Wiley, 2012.

[28] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th. London, UK: Sage Publications, 2013.

[29] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, 2nd ed. Chapman & Hall/CRC, 2013.

[30] M. Alrasheedi, L. F. Capretz, and A. Raza, "Critical success factors for mobile learning: A systematic review and meta-analysis," *Computers in Human Behavior*, vol. 49, pp. 604–620, 2015.

[31] T. Bailey, S. S. Jaggars, and D. Jenkins, *Redesigning America's Community Colleges*. Harvard University Press, 2015.

[32] H. Gangwar, H. Date, and R. Ramaswamy, "Framework for it adoption: A meta-analysis of the toe framework in the context of the information technology innovation," *Journal of Enterprise Information Management*, vol. 28, no. 3, pp. 488–521, 2015.

[33] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, 4th. New York, NY: The Guilford Press, 2015.

[34] B. Roth and J.-E. Gustafsson, "The relation between intelligence and school grades: A meta-analysis," *Intelligence*, vol. 53, pp. 118–137, 2015. DOI: 10.1016/j.intell.2015.09.002.

[35] M. G. R. Alam, A. K. M. Masum, L.-S. Beh, and C. S. Hong, "Critical factors influencing decision to adopt human resource information system (hris) in hospitals," *PLoS ONE*, vol. 11, no. 8, e0160366, 2016.

[36] M. A. Al-Barrak, M. Al-Razgan, and H. S. Al-Khalifa, "Predicting students' final gpa using decision trees: A case study," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 11, pp. 206–211, 2016.

[37] European Union, *Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 (general data protection regulation)*, Official Journal of the European Union, L 119, 1–88, 2016.

[38] S. S. Jaggars and D. Xu, "Online course design features and student performance," *Computers & Education*, vol. 95, pp. 270–284, 2016.

[39] D. A. Canelas, J. L. Hill, and A. Novicki, "Cooperative learning in organic chemistry," *Journal of Chemical Education*, vol. 94, no. 5, pp. 616–622, 2017.

[40] A. F. Hayes, *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 2nd. New York, NY: Guilford Press, 2018.

[41] E. S. Krol, "Association between prerequisites and academic success," *Pharmacy Education Research*, 2019, Prerequisite courses can explain subsequent performance.

[42] J. H. Zar, *Biostatistical Analysis*, 6th. London, UK: Pearson Education, 2019.

[43] Y.-S. Cheng, C.-M. Tsai, and C.-L. Lin, "Integrating the tpack and hot-fit models into a comprehensive framework for mobile learning success in higher education," *Computers & Education*, vol. 156, p. 103 940, 2020.

[44] T. Gonzalez *et al.*, "Influence of covid-19 confinement on students' performance in higher education," *JMIR*, 2020, 22(10): e21359.

[45] J. Jaccard and V. Guilamo-Ramos, *Interaction Effects in Multiple Regression*, 4th. Thousand Oaks, CA: Sage Publications, 2020.

[46] S. Krause-Levy, "Exploring the link between prerequisites and performance," *NSF Publication*, 2020.

[47] E. J. Theobald *et al.*, "Active learning narrows achievement gaps in stem," *Proceedings of the National Academy of Sciences*, vol. 117, no. 12, pp. 6476–6483, 2020.

[48] K. M. Whitcomb *et al.*, "Foundational courses predicting later engineering outcomes," *International Journal of Engineering Education*, vol. 36, no. 6, pp. 1970–1992, 2020.

[49] L. S. Aiken, S. G. West, and R. R. Reno, *Multiple Regression: Testing and Interpreting Interactions*, 2nd. Thousand Oaks, CA: Sage Publications, 2021.

[50] APLU/BTAA, *Gateway course dfw rates and equity gaps*, Learning memo, 2021.

[51] E. Kara, "Class size effects in higher education: Differences across contexts," *Educational Research*, 2021, Effect size around -0.08 in larger classes.

[52] E. Kara *et al.*, "Class size effects in higher education: Differences across disciplines," *Economics of Education Review*, vol. 83, p. 102 138, 2021.

[53] M. Ng, "Class size and student performance in a team-based learning environment," *Journal of Medical Education*, 2021, PMC article.

[54] J. Salmi, "Impact of covid-19 on higher education from an equity perspective," *International Higher Education*, no. 105, pp. 5–7, 2021.

[55] M. A. B. Siddique *et al.*, "Covid-19 and bangladesh education sector," *Asian Journal of Education and Social Studies*, vol. 17, no. 1, pp. 29–40, 2021.

[56] D. Turnbull, R. Chugh, and J. Luck, "Transitioning to e-learning: Hei responses," *Education and Information Technologies*, vol. 26, pp. 6401–6419, 2021.

[57] C. Anfuso *et al.*, "Peer supplemental instruction and dfw patterns," *International Journal of STEM Education*, vol. 9, p. 44, 2022.

[58] T. Azzam *et al.*, "Do learning communities increase first-year retention? rct evidence," *Economics of Education Review*, vol. 89, p. 102 262, 2022.

[59] D. T. Bui, T. T. Nhan, H. T. T. Dang, and T. T. T. Phung, "Online learning experiences of secondary students during covid-19—dataset from vietnam," *Data in Brief*, vol. 45, p. 108 662, 2022.

[60] I. Eteng *et al.*, "Teaching programming in developing countries: A review," *Scientific African*, vol. 16, e01240, 2022.

[61] R. Frei-Landau, "Educational equity amidst the covid-19 pandemic: Exploring the online learning context," *Teaching and Teacher Education*, vol. 105, pp. 103–115, 2022.

[62] M. D. Johnson *et al.*, "Place-based learning community and equity gaps," *CBE–Life Sciences Education*, vol. 21, no. 4, ar65, 2022.

[63] J. Metsämuuronen, "Attenuation-corrected estimators of reliability," *PMC*, 2022.

[64] Z. Yu, "Meta-analyses of differences in blended and traditional learning outcomes," *Frontiers in Psychology*, vol. 13, p. 938 373, 2022.

[65] K. Alom *et al.*, "The covid-19 and online learning process in bangladesh (zoom engagement)," *Heliyon*, vol. 9, no. 8, e19119, 2023.

[66] A. Casanova *et al.*, "Structural barriers to stem retention in higher education," *Frontiers in Education*, 2023.

[67] N. Davidovitch and R. Yavich, “Study group size, motivation and engagement in the digital era,” *Educational Research*, 2023, Motivation differences depend on delivery method.

[68] E. Durand *et al.*, “Technology-enhanced learning in pharmacy programs,” *Exploratory Research in Clinical and Social Pharmacy*, vol. 9, p. 100 206, 2023.

[69] J. Fang, E. Pechenkina, and G. M. Rayner, “Undergraduate learning during covid-19: Remediation insights,” *International Journal of Management Education*, vol. 21, no. 1, p. 100 763, 2023.

[70] E. Goode, T. Roche, and E. Wilson, “Implications of immersive scheduling for student achievement and feedback,” *Studies in Higher Education*, 2023.

[71] M. A. Saleh *et al.*, “A reflection from bangladesh, india, pakistan,” *Frontiers in Education*, vol. 8, p. 10 208 928, 2023.

[72] J. Shao, Y. Chen, X. Wei, X. Li, and Y. Li, “Effects of regulated learning scaffolding on regulation strategies and academic performance: A meta-analysis,” *Frontiers in Psychology*, 2023.

[73] B. Alarifi and S. Song, “Online vs in-person learning in higher education: Effects on student achievement and recommendations for leadership,” *Humanities and Social Sciences Communications*, 2024.

[74] F. Antoniou *et al.*, “Non-linear class size effects (cusp model),” *International Journal of Educational Research Open*, vol. 5, p. 100 217, 2024.

[75] M. Elobaid *et al.*, “Impact of freshman learning communities: A review,” *Frontiers in Education*, vol. 9, p. 1 465 809, 2024.

[76] C. Harper, L. McCormick, and L. Marron, “Face-to-face vs blended learning in higher education: A quantitative analysis of biological science student outcomes,” *International Journal of Educational Technology in Higher Education*, 2024.

[77] M. Nowparvar, N. Soriano, and S. Ozden, “Investigating the impact of learners’ time allocation in immersive simulation-based learning environments,” in *Proceedings of the American Society for Engineering Education*, 2024.

[78] M. A. Rouf *et al.*, *Effectiveness of online education in bangladesh during covid-19*, Publisher: Emerald; Journal abbreviated as PRR, 2024.

[79] A. Shah, V. Agarwal, W. Griswold, and L. Porter, “In-person vs blended learning: An examination of grades, attendance, peer support, competitiveness, and belonging,” in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, 2024.

[80] R. E. Allen *et al.*, “Disruption of covid-19 in university learning: A qualitative study,” *Studies in Higher Education*, 2025.

[81] P. Basu *et al.*, “Twenty-four year study of pltl in general chemistry i,” *PLOS ONE*, vol. 20, no. 7, e0318882, 2025.

[82] Z. Deng *et al.*, “Online education and engagement in the post-covid era,” *Frontiers in Psychology*, vol. 16, p. 1 574 886, 2025.

[83] Financial Times, *Business leaders return to face-to-face learning (executive education trend)*, News article, 2025.

[84] J. M. Robinson *et al.*, “Sense of belonging and dfw reduction,” *Education Sciences*, vol. 15, no. 5, p. 523, 2025.

[85] Saint Louis University, “Impacts of prerequisite proficiency on student performance in managerial economics,” Tech. Rep., 2025, Working paper.

[86] J. Ude, “Enhancing student belonging and academic success through inclusive residential programming in multicultural higher education environments,” *International Journal of Advanced Research in Public Relations*, 2025.

[87] A. S. Wisenöcker *et al.*, “Meta-analysis of academic learning losses post-covid,” *Computers & Education*, vol. 221, p. 105 071, 2025.

Appendix A: Comprehensive Summary of Analyses and Key Findings

This appendix provides a consolidated overview of all hypotheses (H1-H8), including their focus, primary visualizations, major statistical outcomes, interpretations, and actionable recommendations. It serves as a high-level synthesis of the analytical framework presented in Chapter 3.

Table 1: Comprehensive Summary of Hypotheses, Analyses, and Key Findings (Sentence-Level)

RQ/H	Focus / Research Question	Primary Visualization(s)	Main Statistical Result(s)	Interpretation (1 sentence)	Recommendation / Implication (1 sentence)
H1	Delivery mode (Offline vs Hybrid vs Online)	Program-wise mean GPA bars; distribution plots; model summaries	Two-way ANOVA/OLS/Mixed-Effects/GEE: delivery mode highly significant ($p < 0.001$); program main effect significant; *mode*×*program* interaction significant ($p = 0.015$). Online > Hybrid > Offline across many programs.	Learning mode materially affects GPA; magnitude varies by discipline.	Maintain hybrid/online readiness; document grading/assessment changes; use mode-aware supports in programs showing smaller gains.
H2	Courses most impactful on CGPA (by program)	"Top-10 impactful courses" per program; regression scatterplots	Multiple regression shows moderate–high R^2 in several programs (e.g., CSE $\approx$ 0.6549, BIO $\approx$ 0.5801, MAT $\approx$ 0.5159). Many high-impact courses have significant Pearson r (typically $\sim$ 0.40–0.50, $p < 0.001$).	Foundational/high-enrollment courses (e.g., CSE110, MAT110, PHY111, STA201) act as leverage points for CGPA.	Prioritize staffing and pedagogy in these gateways; early-risk flags and targeted tutoring; link course QA to downstream progression.

Continued on next page

Table 1 – continued from previous page

RQ/H	Focus / Research Question	Primary Visualization(s)	Main Statistical Result(s)	Interpretation (1 sentence)	Recommendation / Implication (1 sentence)
H3	Prerequisite → core linkage	Pairwise prereq→core regressions; r, R^2, slope tables	41 pairs analyzed; strongest: CSE321→CSE420 ($r = 0.5962$, $R^2 = 0.3555$). CSE110→CSE111 ($R^2 = 0.2439$), CSE111→CSE220 ($R^2 = 0.2499$). One marginal case: CSE321→CSE410 ($p = 0.057$, not significant at 5%).	Prerequisite mastery is a strong predictor of core success, especially in programming/math chains.	Align LOs/rubrics; add bridge materials/diagnostic refreshers before advanced enrollment; advisor dashboards for prereq-path risks.
H4	Departmental variation in COVID's academic impact	Dept-wise box/violin by COVID period; heatmaps	Department and COVID period are significant predictors across OLS/GEE/Mixed-Effects; Two-way ANOVA shows significant main effects and interaction. During-COVID GPA generally higher; some post-COVID declines (e.g., CSE, SoL).	Pandemic effects were uneven across departments; signs of normalization post-COVID in several units.	DDRF: department-level resilience planning; monitor post-COVID grading rigor; support units showing sharper post-COVID drops.

Continued on next page

Table 1 – continued from previous page

RQ/H	Focus / Research Question	Primary Visualization(s)	Main Statistical Result(s)	Interpretation (1 sentence)	Recommendation / Implication (1 sentence)
H5	Temporal (Pre-, During-, Post-COVID) GPA shifts (overall)	Period-wise bars/box/violin; grade-distribution heatmaps	Mean GPA: Pre $\approx$ 2.85, During $\approx$ 3.40, Post $\approx$ 2.80. Chi-square for grade distributions: χ^2 = 111,577.74, df=22, $p < 0.001$. All models (ANOVA/OLS/Mixed/GEE/Kruskal–Wallis) confirm shifts.	Clear pandemic-era elevation in GPA with partial reversion after; consistent with temporary grading/assessment adjustments.	Preserve effective virtual supports; strengthen online assessment integrity; record policy changes for future comparability.
H6	Class size effect on GPA	Scatter with linear fit; binned boxplots; course/program slices	Overall Pearson $r = 0.0759$ ($p < 0.0001$), slope $\approx$ 0.0045, $R^2 \approx 0.0058$ (very small variance explained). Heterogeneous by program/course (e.g., some positives like MBA ~ 0.17; some slight negatives).	Class size alone weakly relates to GPA; context (discipline/design/instructor) matters more.	Use evidence-based caps where justified; add TAs/structured activities in large sections; explore non-linear and instructor-effects models.
H7	RS vs Main Campus GPA comparison	Mean±SD bars; box/violin; distributions	RS mean $\mu \approx 3.286$ ($\sigma \approx 0.689$, $n = 68{,}503$) vs Main $\mu \approx 2.908$ ($\sigma \approx 1.209$, $n = 822{,}495$). Welch t and Mann–Whitney U: both $p < 0.001$.	RS associates with higher and more consistent outcomes (per-record); per-student repeated-measures models corroborate advantage.	Replicate RS practices (mentoring/structure/engagement) at Main; instrument and track adoption and effects.

Continued on next page

Table 1 – continued from previous page

RQ/H	Focus / Research Question	Primary Visualization(s)	Main Statistical Result(s)	Interpretation (1 sentence)	Recommendation / Implication (1 sentence)
H8	High-failure/withdrawal "hotspots"	Yearly failure-rate trends; program×course heatmaps; top-20 table	Long-run decline (2005→2019), sharp dip in 2019–20, rebound in 2022. Persistent hotspots (e.g., several STEM gateways: examples around 21–25% like CSE220 ∼24.5%, ECE310 ∼22.7%).	Risk is concentrated in repeat-offender gateway courses, not evenly spread.	Early-warning analytics; mandatory supplemental instruction; assessment redesign; enforce prerequisite mastery.
H9	Program and Department-Level Performance Differences	Program-/Department-wise GPA distributions; ANOVA marginal means and radar charts (from H8)	ANOVA and Kruskal–Wallis tests confirm significant GPA variation across programs and departments ($p < 0.001$); post-hoc comparisons show stable rank-order gaps	Meaningful GPA differences exist between programs and departments, reflecting structural influences such as delivery mode, class size, and course design	Implement a Departmental Performance Benchmarking Framework to monitor GPA trends, identify high- and low-performing units, and align instructional and assessment practices across programs.

Note: Visualization references correspond to figures presented in Chapter 3 under each hypothesis. Statistical significance levels are reported at $\alpha = 0.05$ unless otherwise stated.