Business Analytics Digital Assignment
Decoding Learning Patterns:
A Cross-Departmental Analysis of Engagement Metrics and Academic Outcomes at VIT

YUVALAKSHMI M
22MIA1040

1.Introduction

In recent years, educational institutions have witnessed a paradigm shift in learning environments, driven by the increasing integration of digital tools, diverse learning preferences, and evolving pedagogical strategies. At Vellore Institute of Technology (VIT) Chennai , a large and multidisciplinary university, students come from various academic departments, each with unique learning cultures, technological exposures, and instructional methodologies. This research titled *“Decoding Learning Patterns: A Cross-Departmental Analysis of Engagement Metrics and Academic Outcomes at VIT Chennai”* seeks to examine how these differences manifest in terms of student performance, specifically focusing on how learning modes, tools, and departmental contexts influence changes in students’ CGPA. Given the proliferation of hybrid, online, and traditional offline learning modes, it becomes imperative to understand their respective impacts on engagement and academic outcomes. Moreover, students’ use of learning tools such as YouTube, online courses, textbooks, and lecture notes presents an additional dimension that may shape learning effectiveness across disciplines. While some students report higher engagement and performance with flexible, tech-driven modes, others thrive in structured classroom-based environments. This study aims to unravel such patterns through a mixed-method approach, incorporating statistical analyses like Chi-Square tests and Grouping, Summarisation to identify associations between learning-related variables and academic outcomes. By analyzing survey data collected across departments and years of study, the research not only highlights the most influential factors on CGPA change but also explores the perceived impact of learning preferences on academic performance. Ultimately, this investigation contributes to the broader discourse on personalized education, student-centered learning, and evidence-based academic interventions, with the goal of enhancing student success in diverse higher education ecosystems.

2. Research Questions

1. Is there a significant association between preferred learning mode and CGPA change?
2. How does a student's department influence their CGPA change?
3. Do specific learning tools contribute to a positive or negative change in CGPA?
4. Which of the variables (learning mode, department, learning tools) is the strongest predictor of CGPA change?

3. Literature Survey

Several studies have explored the interplay between learning preferences, engagement, and academic outcomes, offering valuable insights that underpin the current research. Awang et al. [1] investigated the relationship between learning style preferences and academic achievement using the VARK model among Malaysian Polytechnic students. A sample of 103 students participated in a descriptive-survey-based study that employed a learning style questionnaire. <u>The findings revealed no significant correlation between students’ preferred learning styles and their academic performance</u>, suggesting that each style holds unique strengths and limitations.

This aligns with the current study's objective to understand whether learning modes have an effect on CGPA change, rather than assuming a direct, one-size-fits-all impact.

Darkwa and Antwi [2] conducted a comparative analysis of classroom and online learning effectiveness at the University of Cape Coast in Ghana during the COVID-19 pandemic. Using a case study method and a paired-sample t-test in SPSS, the researchers analyzed students' performance before and during the shift to online learning. While classroom learning was found to be more effective in terms of pedagogical experience, the difference in academic performance between the two modes was not statistically significant. These findings support the present research's hypothesis that learning mode impacts perceived engagement more than measurable academic outcomes such as CGPA.

Further expanding the role of adaptive technology in learning, El-Sabagh [3] designed an adaptive e-learning environment tailored to students' learning styles and assessed its impact on student engagement. Using a mixed-method approach that included quasi-experimental design and a student engagement scale, the study demonstrated that students in adaptive learning environments exhibited significantly higher engagement than those in traditional settings. This supports the inclusion of tools and learning customization as independent variables in the current study, which also evaluates engagement as a key factor in predicting CGPA change across departments.

Mallillin et al. [4] explored how structural domains of learning (cognitive, affective, psychomotor) and corresponding teaching strategies influence students' academic performance. Conducted with a purposive sample of 30 students, the study utilized a quantitative approach to evaluate how domain-focused teaching strategies affected student analysis, comprehension, and attitudes. Although no statistically significant relationship was found between structural domains and performance, students did show better attention and motivation when strategies aligned with their preferred learning domains. These observations parallel this study's focus on engagement and perceived impact as critical mediating variables between learning tools and academic outcomes.

Lastly, Han and Ellis [5] combined both theory-driven and data-driven approaches to predict academic performance using online learning patterns in a blended course. Self-reported surveys and digital learning trace data were analyzed using hierarchical cluster analysis and sequence clustering, respectively. The study identified deep and surface learning approaches and correlated these with distinct engagement behaviors and academic performance. Results showed that students with higher proportions of active engagement activities achieved better academic outcomes, and importantly, that self-perceptions aligned with observed behaviors. These findings substantiate the mixed-method design used in this research and validate engagement as a predictor of academic success.

4. Variables & Classification

Below are the variables based on survey questions, with one dependent variable (DV) and the rest as independent variables (IVs):

Questionnaire Column	Variable Name	Type	Data Type
Department	`department`	IV	Categorical
Year of Study	`year_of_study`	IV	Numerical

Questionnaire Column	Variable Name	Type	Data Type
Most preferred learning mode	`preferred_learning_mode`	IV	Categorical
Which learning tools do you use the most for studying?	`learning_tools_used`	IV	Categorical
How engaged do you feel during your preferred learning mode?	`engagement_level`	IV	Numerical (Scale)
Your level of focus and attentiveness during learning sessions?	`focus_level`	IV	Numerical (Scale)
How has your CGPA changed since using your preferred learning mode?	`cgpa_change`	DV	Categorical
To what extent do you believe your learning preferences impact your academic performance?	`perceived_impact`	IV	Numerical (Scale)
What is the biggest advantage of your preferred learning mode?	`learning_advantage`	IV	Categorical
What is the biggest challenge in your preferred learning mode?	`learning_challenge`	IV	Categorical
Which learning tool do you think will be most useful in the future?	`future_learning_tool`	IV	Categorical

Regarding the categorical/numerical classification of variables:

- **Categorical Variables**: `department`, `year_of_study`, `preferred_learning_mode`, `learning_tools_used`, `cgpa_change`, `learning_advantage`, `focus_level`, `learning_challenge`, `future_learning_tool`
- **Numerical Variables**: `engagement_level`, `perceived_impact`

5. Research Methodology

5.1 Research Design Incorporating statistical techniques to analyze the relationships between learning mode, department, and learning tools on CGPA change.

5.2 Sampling Method & Data Collection

Although a stratified random sampling method was initially intended, the final distribution shows uneven representation across years and departments. Therefore, the actual sampling approach leans toward **convenience sampling**, which may limit the generalizability of the results due to sampling bias. Data was collected through an online questionnaire using fillout website.

Dataset:

department	year_of_study	preferred_learning_mode	learning_tools_used	engagement_level	focus_level	cgpa_change	perceived_impact	learning_advantage	learning_challenge	future_learning_tool
AI	Third Year	Online	YouTube	4	Moderately engaged	No change	83	Better understanding of conce	Lack of motivation, Intern	AI-powered tutoring tools
AI	Third Year	Online	YouTube,Online Courses -	4	Occasionally distracted	Slight increase (0.1 - 0.4)	84	Better understanding of conce	Lack of motivation	AI-powered tutoring tools
SOFTWARE ENGINEERIN	Third Year	Online	YouTube,Online Courses -	4	Moderately engaged	Slight increase (0.1 - 0.4)	79	Flexibility, Better understandin	Lack of motivation, Poor i	AI-powered tutoring tools
SOFTWARE ENGINEERIN	Third Year	Online	YouTube,Online Courses -	4	Moderately engaged	Slight increase (0.1 - 0.4)	89	Flexibility, Better understandin	Lack of motivation	AI-powered tutoring tools
SOFTWARE ENGINEERIN	Third Year	Offline	Lecture Notes,Textbooks	4	Moderately engaged	Slight increase (0.1 - 0.4)	100	Better understanding of conce	Poor interaction with instr	YouTube

5.3 Data Analysis Techniques

- **Grouping and Summarization:** To analyze CGPA trends across departments based on preferred learning mode and learning tools.
- **Chi-Square Test:** To check for statistical associations between categorical variables and CGPA change.
- **Pie Chart and Bar Chart Analysis:** To visually represent CGPA variations and detect outliers, patterns, and trends.

6. Questionnaire

FORM LINK: https://yuva.fillout.com/learning_preference_survey

Demographic Information Questionnaire

1. **Name:** __________________________ (First and Last Name)
2. **Department:**
    - CSE
    - AI
    - EEE
    - BUSINESS ANALYTICS
    - SOFTWARE ENGINEERING
    - LAW
    - BCOM
    - BSc
3. **Year of Study:**
    - First Year
    - Second Year
    - Third Year
    - Fourth Year
    - Other: _______________ (Please specify)

Learning Preferences Questionnaire

1. **What is your most preferred learning mode?**
    - Offline
    - Online
    - Hybrid
2. **Which learning tools do you use the most for studying?**
   __________________________ (Please specify)
    - YouTube
    - Lecture Notes
    - Online Courses
    - Study groups
    - Textbooks
3. **What is your preferred learning schedule?**
    - Morning
    - Afternoon
    - Evening
    - Flexible

Engagement Levels Questionnaire

How many hours per day do you dedicate to studying through your preferred mode? ___________________________

1. **How engaged do you feel during your preferred learning mode?** Please rate your engagement on a scale of 1 to 5, where 1 is the lowest engagement and 5 is the highest.
   - 1
   - 2
   - 3
   - 4
   - 5

2. **What is your level of focus and attentiveness during learning sessions?**
   - Highly engaged
   - Moderately engaged
   - Occasionally distracted
   - Frequently distracted
   - Not engaged at all

Academic Performance and Self-Assessment Questionnaire

1. **How has your CGPA changed since using your preferred learning mode?**
   - Increased significantly (+0.5 or more)
   - Slight increase (0.1 - 0.4)
   - No change
   - Slight decrease (-0.1 to -0.4)
   - Decreased significantly (-0.5 or more)
2. **To what extent do you believe your learning preferences impact your academic performance?**

   Indicate your level on a scale from 1 to 100, where 1 means "Not at all" and 100 means "To a very great extent".

   ________________________________ (Please enter a number between 1 and 100)

Questionnaire on Learning Preferences and Future Tools

Section 1: Preferred Learning Mode

1. **What is the biggest advantage of your preferred learning mode?**
   - (a) Flexibility
   - (b) Better understanding of concepts
   - (c) Higher engagement
   - (d) More convenient resources
2. **What is the biggest challenge in your preferred learning mode?**
   - (a) Lack of motivation

- (b) Difficulty in understanding concepts
- (c) Poor interaction with instructors
- (d) Internet / connectivity issues

Section 2: Future of Learning Tools

3. **Which learning tool do you think will be most useful in the future?**
   - (a) YouTube
   - (b) Coursera / Udemy
   - (c) Lecture notes
   - (d) Live lectures
   - (e) AI-powered tutoring tool

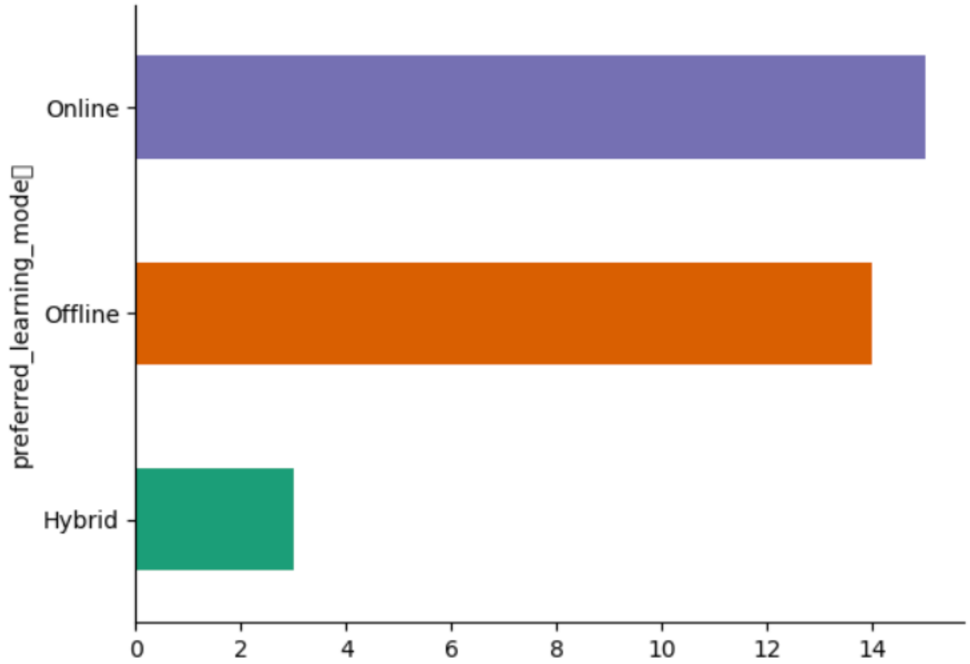
7. Checking for Missing values and Handle NaNs

```
[ ] # Check for missing values
    print(df.isnull().sum())

    # Fill missing categorical values with mode
    for col in df.select_dtypes(include=["object"]).columns:
        df[col].fillna(df[col].mode()[0], inplace=True)

    # Verify again
    print(df.isnull().sum())
```

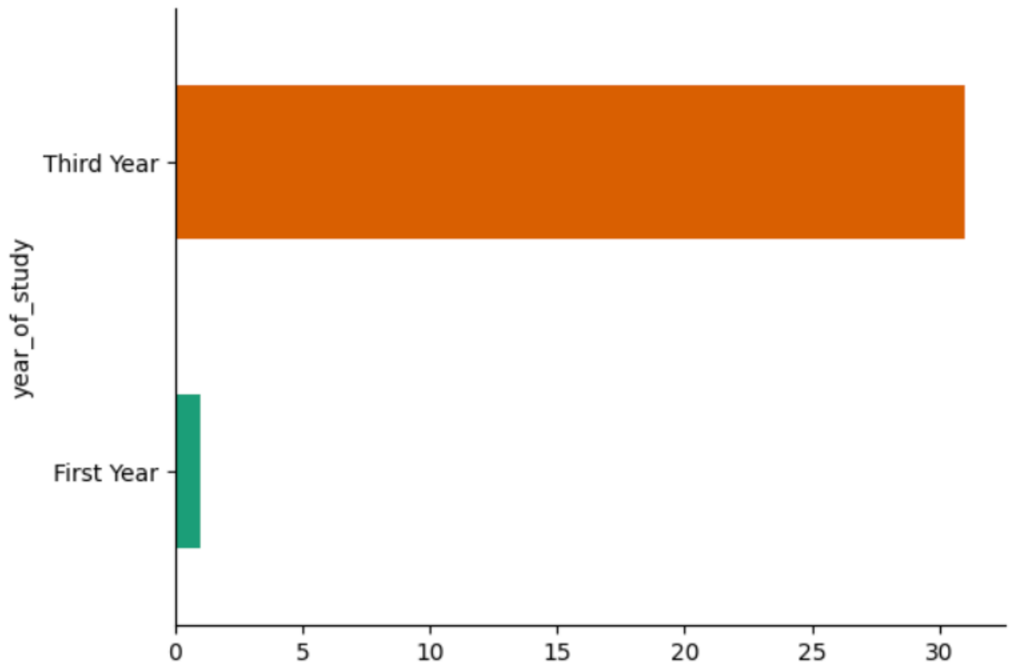
```
department               0
preferred_learning_mode  0
learning_tools_used      0
engagement_level         0
focus_level              0
cgpa_change              0
dtype: int64
department               0
preferred_learning_mode  0
learning_tools_used      0
engagement_level         0
focus_level              0
cgpa_change              0
dtype: int64
```

All data columns are now **complete and clean**, which is critical before performing any analysis or model training. Using the mode ensures **consistency** in categorical data without introducing new categories or arbitrary values.

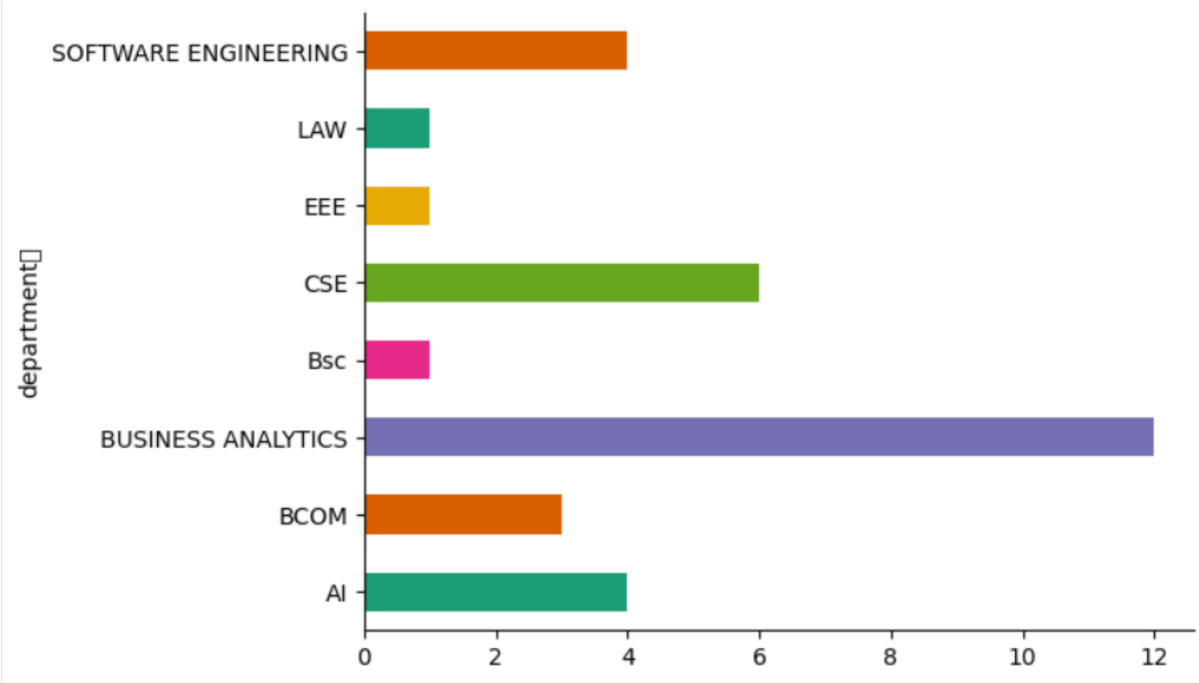
8. Pie Chart and Bar Chart Analysis



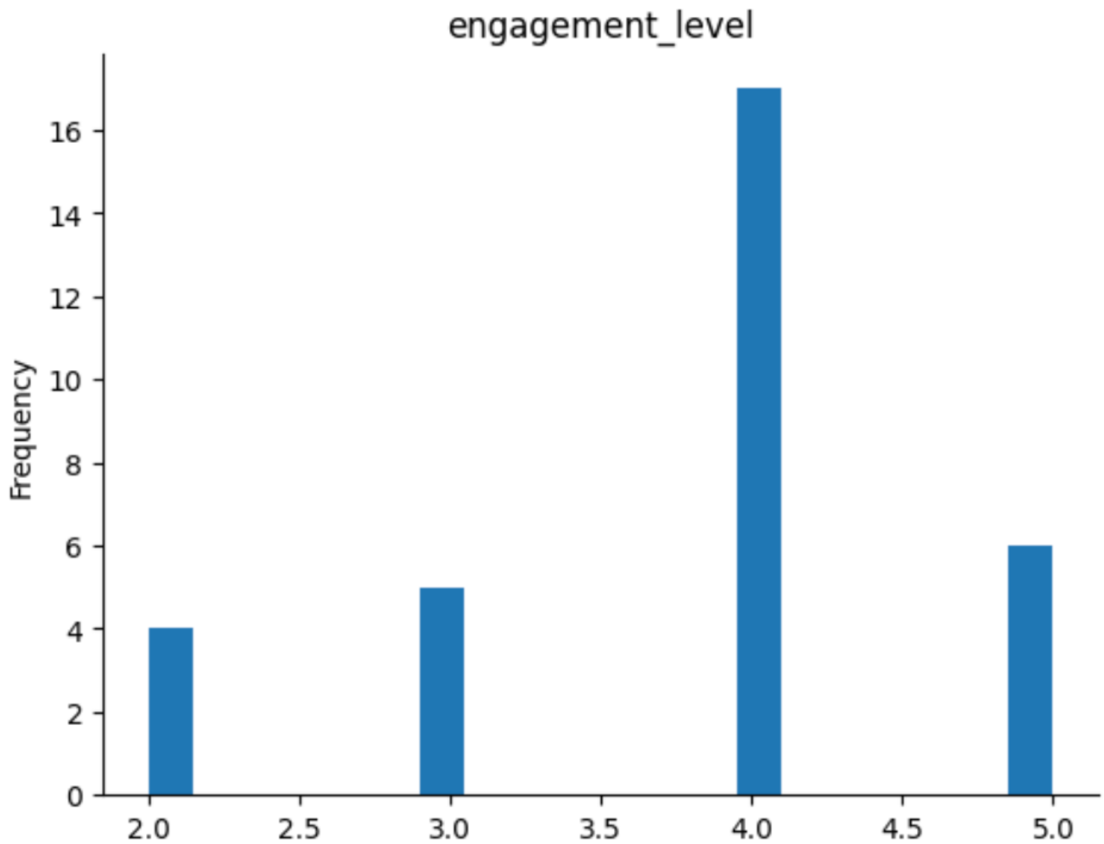
Interpretation: Indicates that most students prefer either online or offline learning modes, with hybrid learning being the least favored.



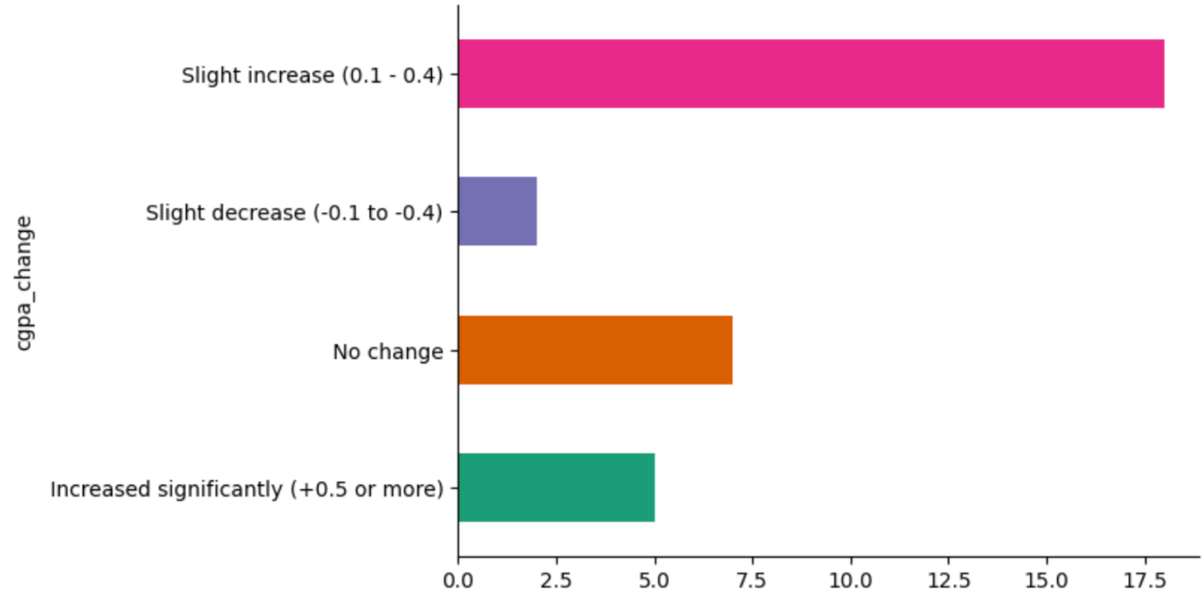
Interpretation: Shows that the majority of survey respondents are third-year students, while first-year students are minimally represented.



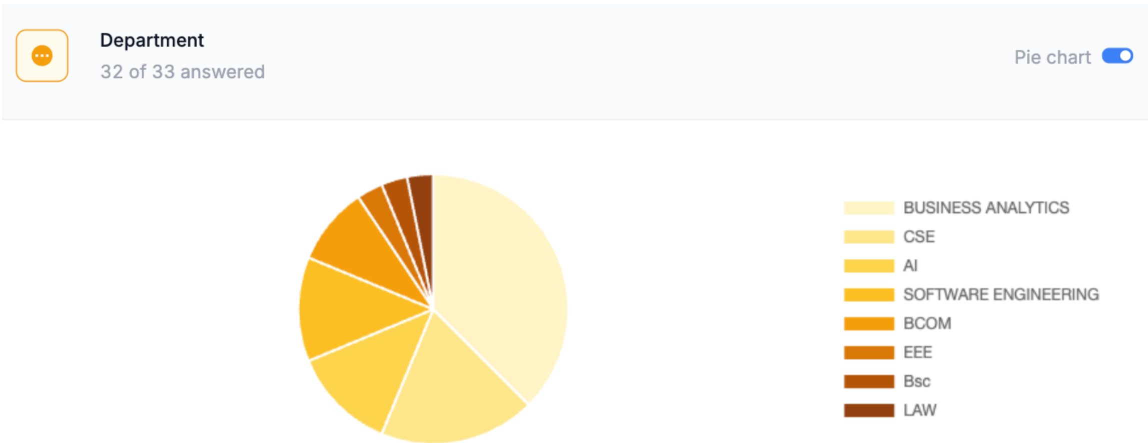
Interpretation: The **Business Analytics** department has the highest student representation, while departments like EEE and LAW have the least.



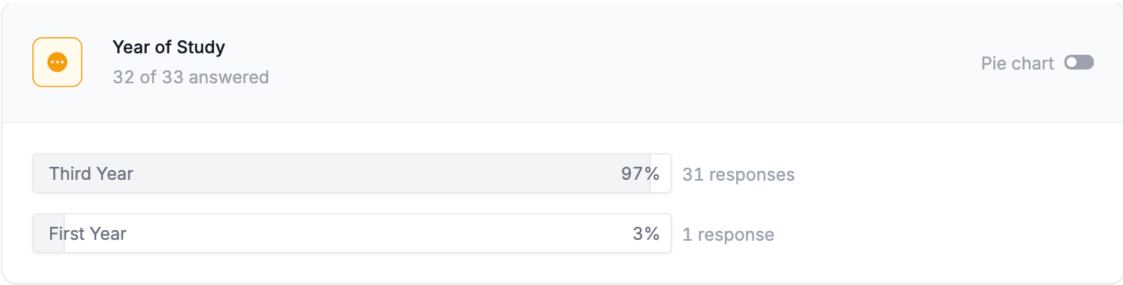
Interpretation: Most students rated their **engagement level as 4**, indicating relatively high engagement across their preferred learning mode.

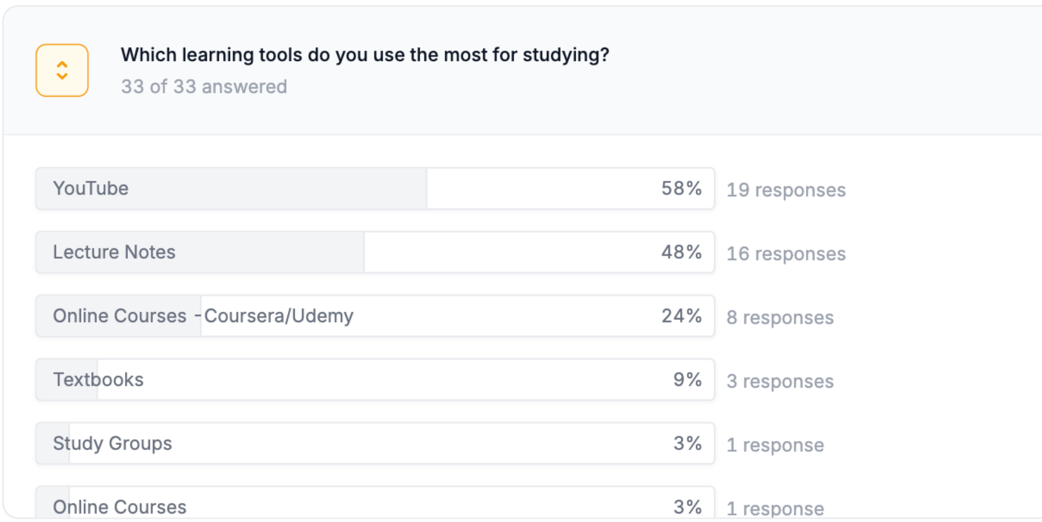


Interpretation: The majority of students experienced a **slight GPA increase**, followed by those reporting **no change**.

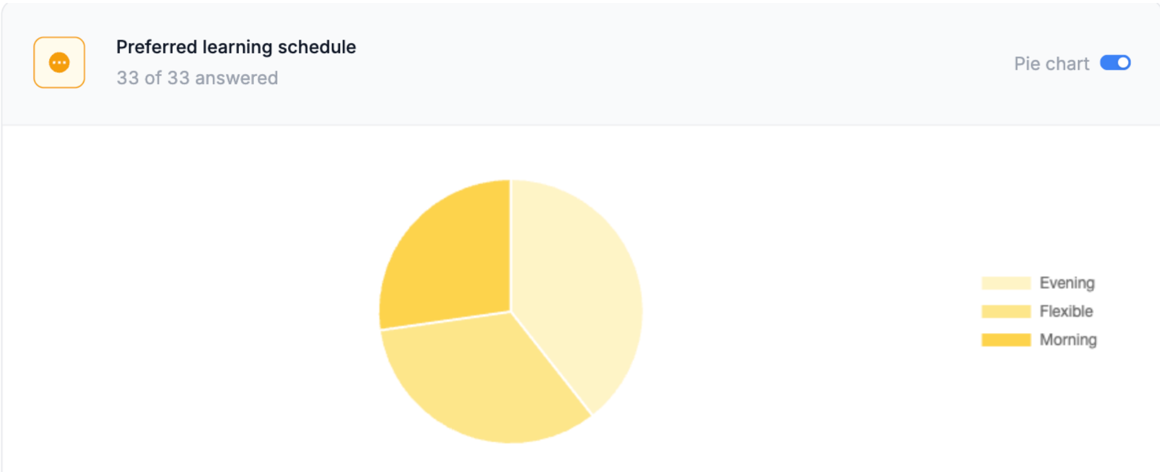


Interpretation: Business Analytics and CSE dominate the departmental composition, suggesting they are the most populous.

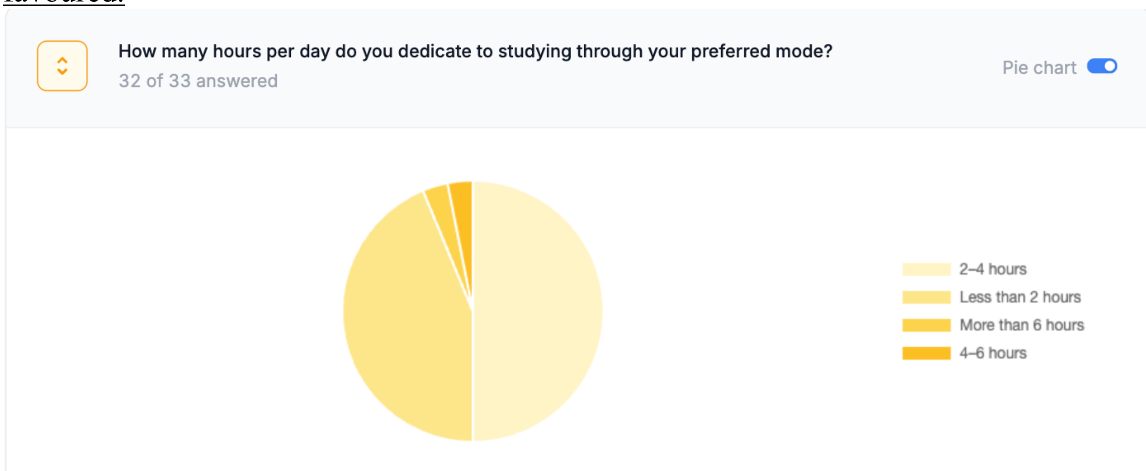




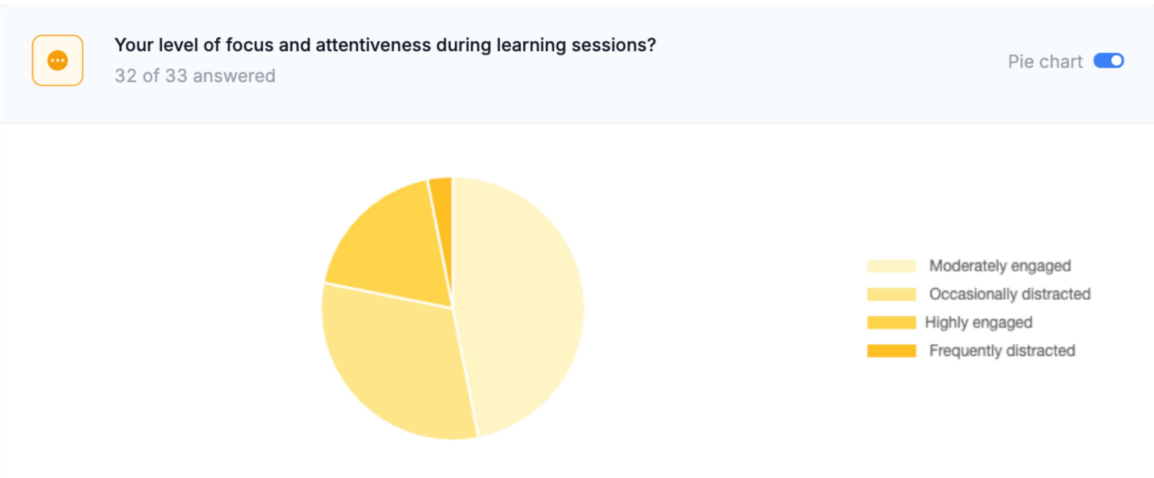
YouTube (58%) and Lecture Notes (48%) are the most commonly used learning resources among students.



Evening is the most preferred study schedule, followed by **flexible** timing; morning is the least favoured.

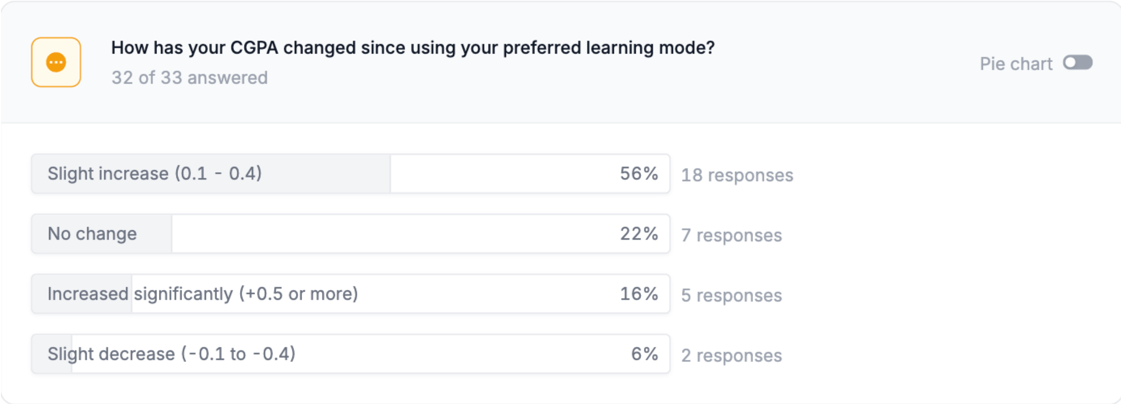


Most students study for **2–4 hours daily**, and a significant number report an engagement level of **4**, showing moderately high engagement in preferred learning modes.



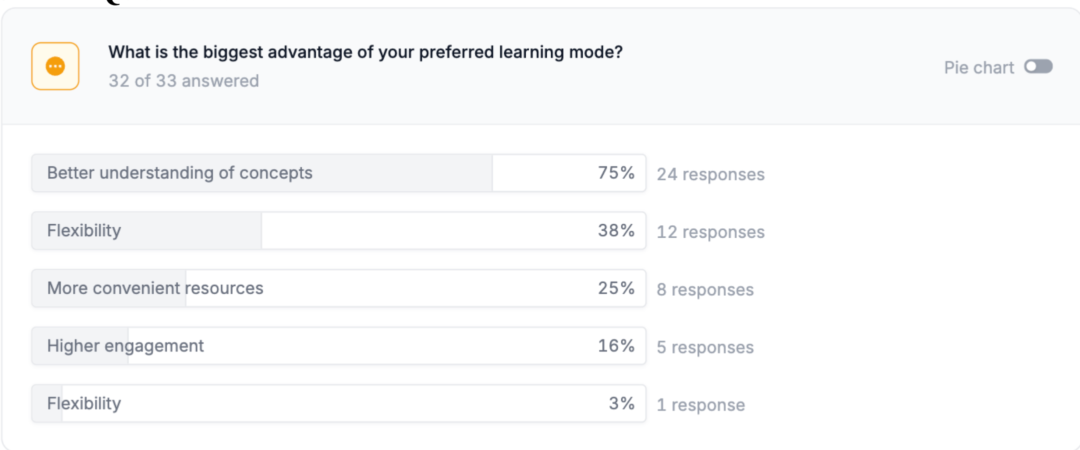
- **Moderately engaged** and **occasionally distracted** dominate.
- A smaller chunk reports being **highly engaged**.
- Very few are **frequently distracted**.

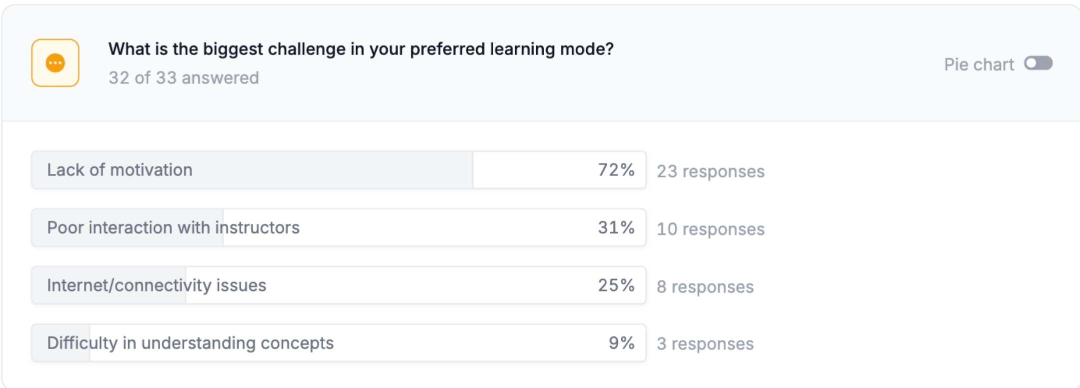
Interpretation: While distractions are present, most students stay reasonably focused during study sessions.

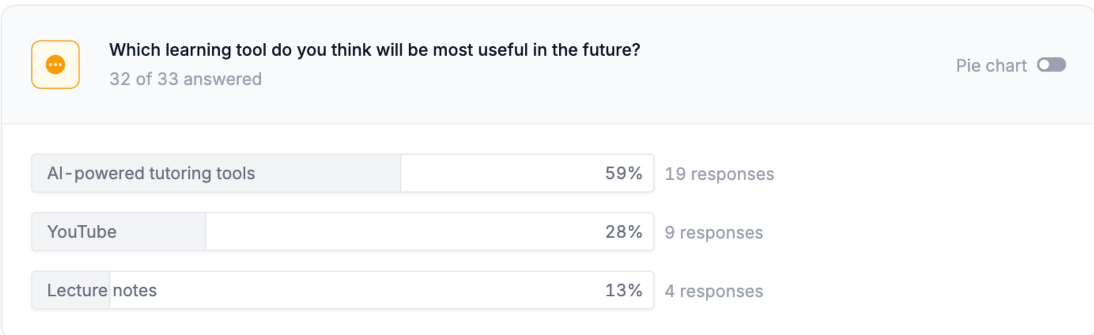


The preferred learning modes have **positively impacted academic performance** for most students.

#Extra Questions







9. Grouping and Summarising mean CGPA across departments

```python
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
from scipy.stats import f_oneway


# Function to round to valid CGPA values
def round_to_valid_cgpa(value):
    valid_values = [0, 1, -0.5, 0.5]
    closest_value = min(valid_values, key=lambda x: abs(x - value))
    return closest_value

# Preprocessing: Clean department and other categorical columns
df['department'] = df['department'].str.strip().str.upper()  # Clean department names
df['preferred_learning_mode'] = df['preferred_learning_mode'].str.strip().str.upper()  # Clean learning mode
df['learning_tools_used'] = df['learning_tools_used'].str.strip().str.upper()  # Clean learning tools

# Group by department and preferred learning mode, calculating average CGPA change and rounding
learning_mode_efficiency = df.groupby(["department", "preferred_learning_mode"])["cgpa_change"].apply(lambda x: pd.to_numeric(x, errors=
learning_mode_efficiency["cgpa_change"] = learning_mode_efficiency["cgpa_change"].apply(round_to_valid_cgpa)

# Further aggregate to have one row per department and learning mode
learning_mode_efficiency = learning_mode_efficiency.groupby(["department", "preferred_learning_mode"])["cgpa_change"].mean().reset_index
learning_mode_efficiency["cgpa_change"] = learning_mode_efficiency["cgpa_change"].apply(round_to_valid_cgpa) # Round again after aggrega


# Group by department and learning tool, calculating average CGPA change and rounding
learning_tool_efficiency = df.groupby(["department", "learning_tools_used"])["cgpa_change"].apply(lambda x: pd.to_numeric(x, errors='co
learning_tool_efficiency["cgpa_change"] = learning_tool_efficiency["cgpa_change"].apply(round_to_valid_cgpa)

# Further aggregate to have one row per department and learning tool
learning_tool_efficiency = learning_tool_efficiency.groupby(["department", "learning_tools_used"])["cgpa_change"].mean().reset_index()
learning_tool_efficiency["cgpa_change"] = learning_tool_efficiency["cgpa_change"].apply(round_to_valid_cgpa) # Round again after aggrega

# Display department-wise preferred learning mode and CGPA change
print("Department-wise Preferred Learning Mode and CGPA Change:")
for department in learning_mode_efficiency["department"].unique():
    department_data = learning_mode_efficiency[learning_mode_efficiency["department"] == department]
    print(f"\nDepartment: {department}")
    for index, row in department_data.iterrows():
        print(f"  - {row['preferred_learning_mode']}: {row['cgpa_change']}")

# Display department-wise learning tools used and CGPA change
print("\nDepartment-wise Learning Tools Used and CGPA Change:")
for department in learning_tool_efficiency["department"].unique():
    department_data = learning_tool_efficiency[learning_tool_efficiency["department"] == department]
    print(f"\nDepartment: {department}")
    for index, row in department_data.iterrows():
        print(f"  - {row['learning_tools_used']}: {row['cgpa_change']}")
```

```
Department-wise Preferred Learning Mode and CGPA Change:

Department: AI
  - OFFLINE: -0.5
  - ONLINE: 0.5

Department: BCOM
  - OFFLINE: 1.0
  - ONLINE: 0.0

Department: BSC
  - HYBRID: 0.5

Department: BUSINESS ANALYTICS
  - HYBRID: 1.0
  - OFFLINE: 0.0
  - ONLINE: 0.5

Department: CSE
  - OFFLINE: 0.5
  - ONLINE: 0.5

Department: EEE
  - OFFLINE: 1.0

Department: LAW
  - HYBRID: 0.0

Department: SOFTWARE ENGINEERING
  - OFFLINE: 1.0
  - ONLINE: 1.0
```

```
Department-wise Learning Tools Used and CGPA Change:

Department: AI
  - YOUTUBE: 0.0
  - YOUTUBE,LECTURE NOTES: -0.5
  - YOUTUBE,ONLINE COURSES -COURSERA/UDEMY: 1.0

Department: BCOM
  - TEXTBOOKS: 0.5
  - YOUTUBE: 1.0
  - YOUTUBE,LECTURE NOTES: 0.0

Department: BSC
  - LECTURE NOTES: 0.5

Department: BUSINESS ANALYTICS
  - LECTURE NOTES: 0.5
  - ONLINE COURSES -COURSERA/UDEMY,STUDY GROUPS,LECTURE NOTES: 1.0
  - TEXTBOOKS: 0.0
  - YOUTUBE: 0.0
  - YOUTUBE,LECTURE NOTES: 0.5

Department: CSE
  - LECTURE NOTES: -0.5
  - LECTURE NOTES,ONLINE COURSES -COURSERA/UDEMY: 1.0
  - YOUTUBE: 1.0
  - YOUTUBE,LECTURE NOTES: 1.0
  - YOUTUBE,ONLINE COURSES -COURSERA/UDEMY: 0.0

Department: EEE
  - YOUTUBE: 1.0

Department: LAW
  - ONLINE COURSES -COURSERA/UDEMY: 0.0

Department: SOFTWARE ENGINEERING
  - LECTURE NOTES: 1.0
  - LECTURE NOTES,TEXTBOOKS: 1.0
  - YOUTUBE,ONLINE COURSES -COURSERA/UDEMY: 1.0
```

Interpretation:

The learning_mode_efficiency output shows, for example, in the AI department, the average CGPA change is -0.5 for OFFLINE mode and 0.5 for ONLINE mode.

The learning_tool_efficiency output shows, for instance, in the BCOM department, using TEXTBOOKS results in an average CGPA change of 0.5 while using YOUTUBE leads to an average CGPA change of 1.0.

10. Chi Square Test:-

```
[ ] from scipy.stats import chi2_contingency

    # Perform Chi-Square test for each IV against CGPA Change
    for col in ["preferred_learning_mode", "department","learning_tools_used"]:
        contingency_table = pd.crosstab(df[col], df["cgpa_change"])
        chi2, p, dof, expected = chi2_contingency(contingency_table)
        print(f"Chi-Square Test for {col} vs CGPA Change:")
        print(f"Chi2 = {chi2}, p-value = {p}\n")
```

```
Chi-Square Test for preferred_learning_mode vs CGPA Change:
Chi2 = 10.006107331821617, p-value = 0.12439506199736193

Chi-Square Test for department vs CGPA Change:
Chi2 = 29.56190476190476, p-value = 0.1011514448126585

Chi-Square Test for learning_tools_used vs CGPA Change:
Chi2 = 20.011791383219954, p-value = 0.6961053943061429
```

Inference from the Chi-Square Test Results:-
The ChiSquare test checks whether there is a significant relationship between the independent *variables (IVs) and the dependent variable (CGPA Change).*

1. Preferred Learning Mode vs CGPA Change

- **Null Hypothesis (H_0):** There is no significant association between preferred learning mode and CGPA change.
- **Alternative Hypothesis (H_1):** There is a significant association between preferred learning mode and CGPA change.

Test Results:
$Chi^2 = 10.006$, p-value = 0.1244

Since p-value > 0.05, we fail to reject the null hypothesis.
Interpretation: There is no statistically significant relationship between preferred learning mode and CGPA change.

2. Department vs CGPA Change

- **Null Hypothesis (H_0):** There is no significant association between department and CGPA change.

- **Alternative Hypothesis (H_1):** There is a significant association between department and CGPA change.

Test Results:
$Chi^2 = 29.562$, p-value = 0.1012

Since p-value > 0.05, we fail to reject the null hypothesis.
Interpretation: Department does not have a statistically significant impact on CGPA change.

3. Learning Tools Used vs CGPA Change

- **Null Hypothesis (H_0):** There is no significant association between learning tools used and CGPA change.
- **Alternative Hypothesis (H_1):** There is a significant association between learning tools used and CGPA change.

Test Results:
$Chi^2 = 20.012$, p-value = 0.6961

Since p-value > 0.05, we fail to reject the null hypothesis.
Interpretation: Learning tools used do not significantly affect CGPA change.

Based on the Chi-Square test results, all p-values for the independent variables **Preferred Learning Mode**, **Department**, and **Learning Tools Used** were greater than 0.05. This indicates that none of these variables have a statistically significant impact on CGPA change.

Conclusion:
This study explored the relationship between students' preferred learning modes, departmental backgrounds, and the tools they use, in relation to changes in their CGPA. Despite expectations that these variables might significantly influence academic performance, statistical analyses including Chi-Square tests revealed no significant associations. Whether students preferred online, offline, or hybrid learning; came from different departments; or used various learning tools like YouTube or textbooks—none of these factors showed a consistent impact on CGPA change.

These findings suggest that academic outcomes are likely shaped by more complex and individualized factors beyond surface-level preferences or tools. While learning modes and tools may enhance subjective learning experiences, they are not standalone predictors of academic success. Future research could delve deeper into personal habits, motivation, or instructor quality to uncover more influential drivers of performance.

References

[1] H. Awang, N. Abd Samad, N. S. Mohd Faiz, R. Roddin and J. D. Kankia, "Relationship between the Learning Styles Preferences and Academic Achievement," *IOP Conference Series: Materials Science and Engineering*, vol. 226, pp. 012193, 2017. doi: 10.1088/1757-899X/226/1/012193

[2] B. F. Darkwa and S. Antwi, "From Classroom to Online: Comparing the Effectiveness and Student Academic Performance of Classroom Learning and Online Learning," *Open Access Library Journal*, vol. 8, p. e7597, Jul. 2021. doi: 10.4236/oalib.1107597

[3] H. A. El-Sabagh, "Adaptive e-learning environment based on learning styles and its impact on development students' engagement," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 53, 2021. doi: https://doi.org/10.1186/s41239-021-00289-4

[4] L. L. D. Mallillin, J. C. Cabaluna, R. D. Laurel, P. A. C. Arroyo, T. M. Señoron Jr and J. B. Mallillin, "Structural Domain of Learning and Teaching Strategies in the Academic Performance of Students," *European Journal of Education Studies*, vol. 8, no. 9, 2021. doi: 10.46827/ejes.v8i9.3902

[5] F. Han and R. A. Ellis, "Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course," *Educational Technology & Society*, vol. 24, no. 1, pp. 191-204, Jan. 2021. Available: https://www.jstor.org/stable/10.2307/26977867