LPA Analysis:

Through a literature review, the common determiners of effort in a behavioural framework are time and effort. As we do not have cohort information, we will simulate the difference in engagement across cohorts (over time) by splitting the dataset in half based on the time of the first interaction.

Methodology

Defining Variables

This study's definition of student engagement is based on Berman and Artino's conceptualization, which views behavioural engagement as observable actions, such as time spent on task and answer accuracy.

- **Time**: Time spent on a task is measured in seconds spent on each page of the ophthalmology module. The total time per page will be used as a metric to assess the level of engagement. This helps to identify whether students spend sufficient time interacting with course content or are rushing through materials without adequate attention.
- **Effort**: Effort is measured by scoring student responses to writing prompts within the module. Each student is required to provide written responses to progress through the content. The quality of the responses will be evaluated using Sentiment Analysis via a Natural Language Processing (NLP) model. This model will categorize responses based on their sentiment, from neutral to positive. Additional metrics, such as complexity, will also be used to measure the depth of students' responses.
    - The response score represents the level of effort a student has invested in the task, with higher scores reflecting more substantial or thoughtful contributions (e.g., complete sentences) and lower scores indicating minimal engagement (e.g., single words or letters).

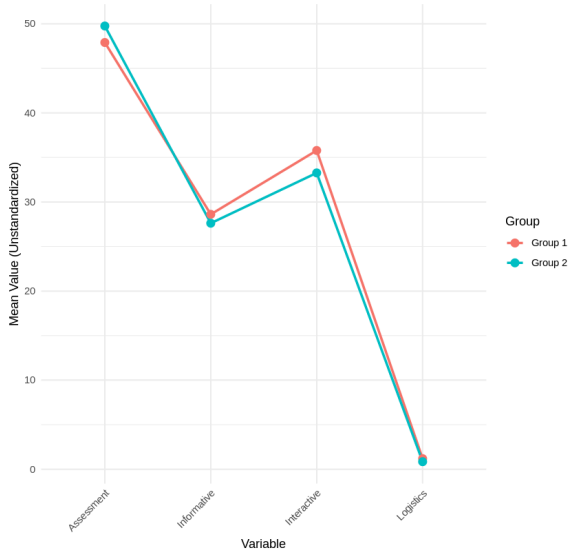
Latent Profile Analysis (LPA)

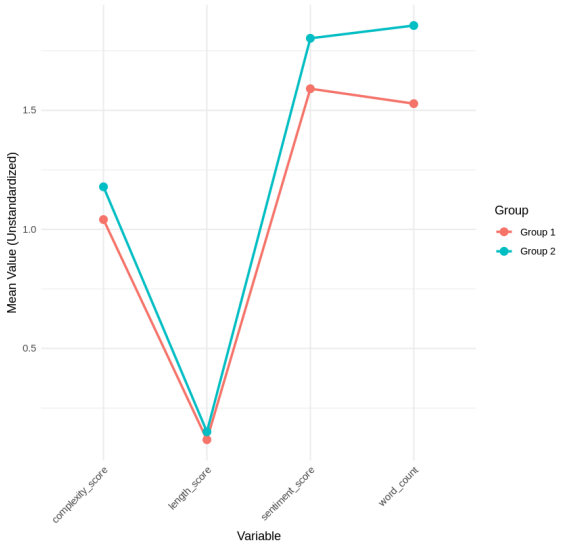
Latent Profile Analysis (LPA) was performed to identify distinct engagement profiles within the two independent groups of students.

1. **Time-Based Grouping**:

    - The data were split into two independent groups based on the time of interaction:
        - Group 1: Data from November 1, 2022 to October 31, 2023
        - Group 2: Data from November 1, 2023 to November 1, 2024
2. **Variable Selection and Data Structuring**:
    - Two main data frames were created from the xAPI data:
        - **Time**: The total time spent per page was calculated by summing each student's time (in seconds) spent on each page. The sum was used rather than the mean to avoid the influence of students who may click through slides quickly without engaging with the content.
        - **Text Scores**: The quality of student-written responses was scored using an NLP model. Responses were rated on a scale from 0 to 10 based on sentiment (ranging from neutral to positive) and complexity.
3. **Categorization of Module Pages**:
    - Module pages were categorized according to the type of engagement they prompted:
        - **Assessment**: Pages that prompt students to respond.
        - **Informative**: Pages that provide information about the case study or patient.
        - **Interactive**: Pages that test students' knowledge through review and practice exercises.
        - **Logistics**: Pages providing basic information about the module or surveys.
4. **Outlier Detection**:
    - To clean the data, the Interquartile Range (IQR) method was used to remove outliers. Initially, a rigid IQR formula was applied, but it was adjusted to expand the appropriate range based on the dataset's distribution. This adjustment ensured that outliers did not disproportionately affect the analysis.
5. **LPA Process**: The LPA was used to classify students into high, moderate, or low engagement profiles based on their engagement patterns. This allows for a deeper understanding of the distribution of engagement levels within the two groups and how these profiles differ across the two time periods.
6. **Statistical Analysis:** Statistical comparisons through the Mann-Whitney U-Test were conducted to assess the differences between the engagement profiles (high, moderate, low) across the two groups of students. These analyses will help determine if the two groups significantly differ in engagement patterns.

Preliminary Visualization



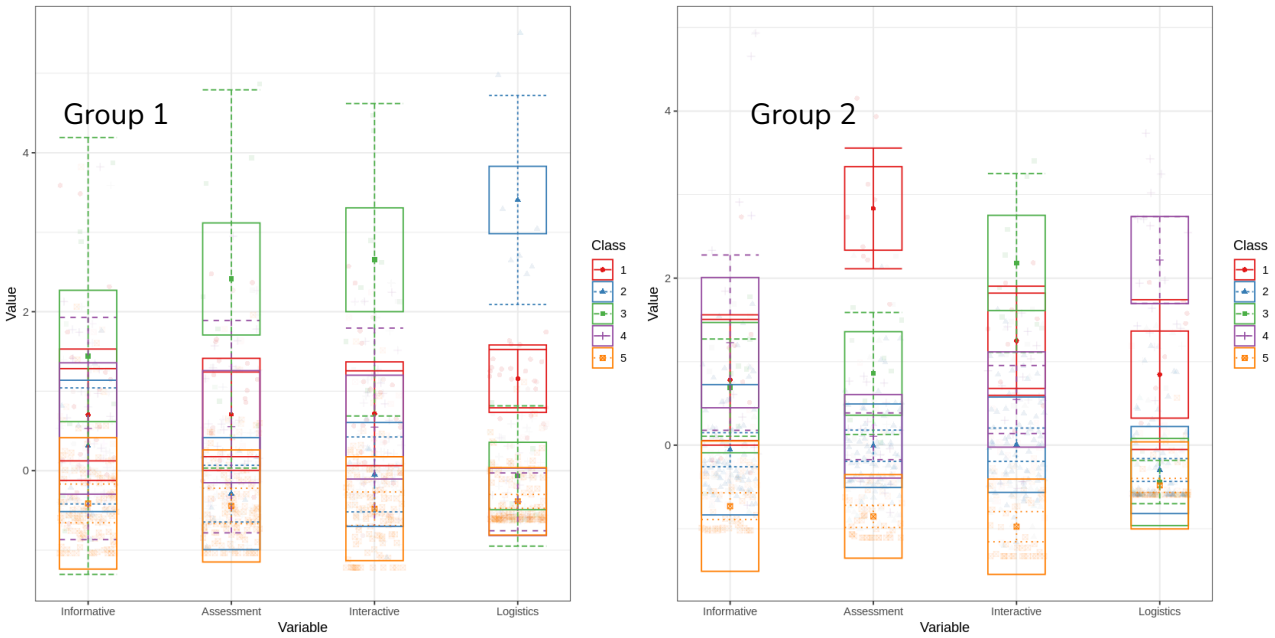


Comparison of mean values between groups (Time) Comparison of mean values between groups (Text scores)

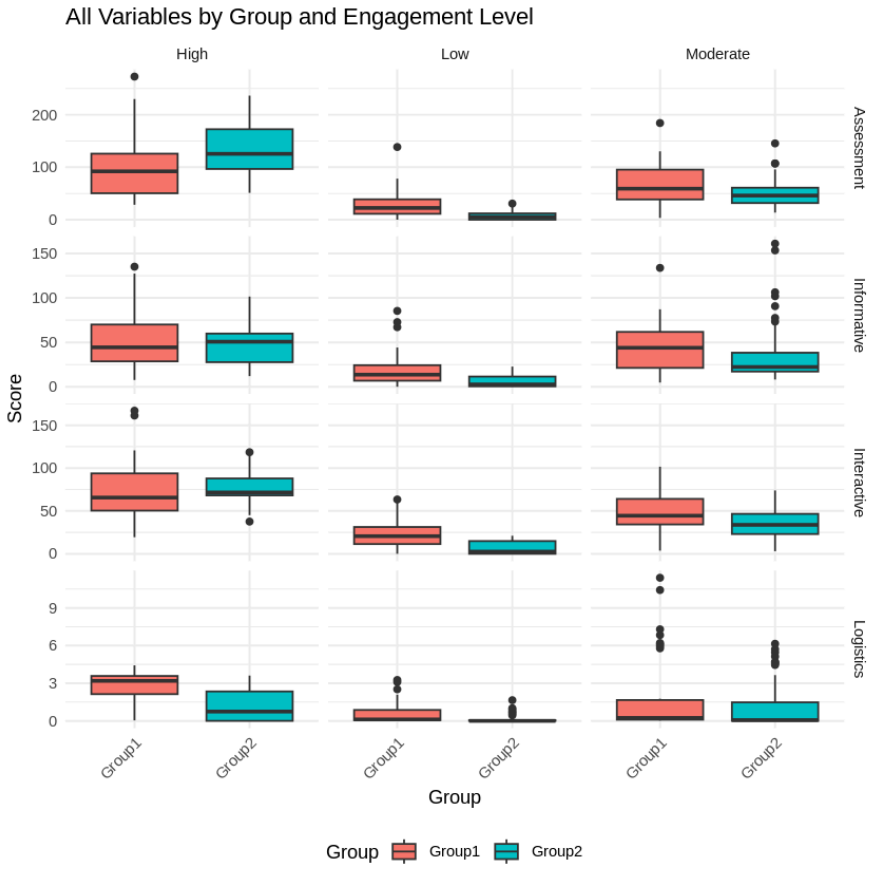
Both charts indicate that the two groups are quite similar in performance, with the latter group (group 2) showing slightly better performance in both factors.

Time Spent per Page Analysis

Profiles



A 5 profile solution was the best fit for both groups. I then grouped profiles into High, Moderate and Low engagement profiles based on their distance from the mean. This is to ensure that we are not accidentally comparing high and moderate across the two time groups.



	Variable	Profile	W	p_value	effect_size	n_group1	n_group2	interpretation
W3	Assessment	High	165.0	0.025	0.320	32	17	Medium effect
W	Informative	High	283.0	0.825	0.032	32	17	Negligible effect
W6	Interactive	High	225.0	0.329	0.140	32	17	Small effect
W9	Logistics	High	440.0	0.000	0.503	32	17	Large effect
W5	Assessment	Low	3160.0	0.000	0.432	111	36	Medium effect
W2	Informative	Low	2997.5	0.000	0.371	111	36	Medium effect
W8	Interactive	Low	3209.0	0.000	0.451	111	36	Medium effect
W11	Logistics	Low	3300.5	0.000	0.485	111	36	Medium effect
W4	Assessment	Moderate	1912.0	0.033	0.192	34	90	Small effect
W1	Informative	Moderate	2018.0	0.006	0.245	34	90	Small effect
W7	Interactive	Moderate	2090.0	0.002	0.281	34	90	Small effect
W10	Logistics	Moderate	2029.0	0.005	0.254	34	90	Small effect

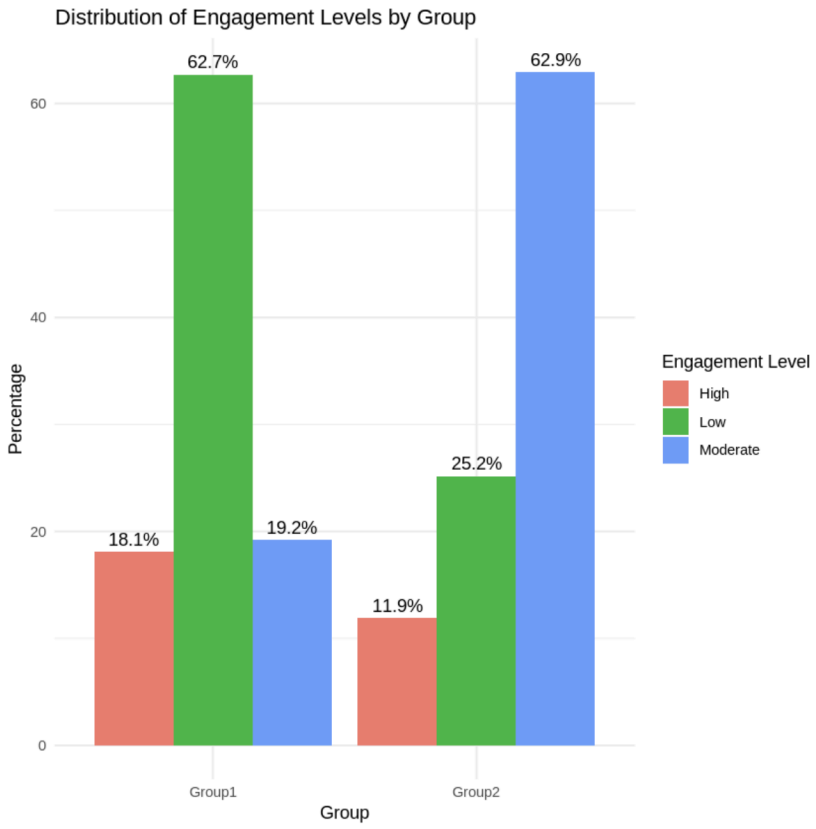
For the high-profile assessment comparison, there was a statistically significant medium effect size, where group two showed higher values in the Assessment category.

Moderate engagement profiles showed a small effect size of statistical significance for all variables. (p = 0.002, 0.005, 0.006, 0.033)

Low engagement profiles showed a medium effect size of high statistical significance (p = 0.000) for all variables.

We can reject the null hypothesis as the results are unlikely to be due to chance. In the moderate and low engagement profiles, group two spends less time across all categories, indicating less engagement As for the high profiles, there is no significant difference between the two groups except for the Assessment category, where students in Group 2 show more time invested and thus higher engagement compared to Group 1. However, this could be due to the small group sizes (n<30 for both groups in high-engagement profile) which raises questions about the power of the mean comparisons.

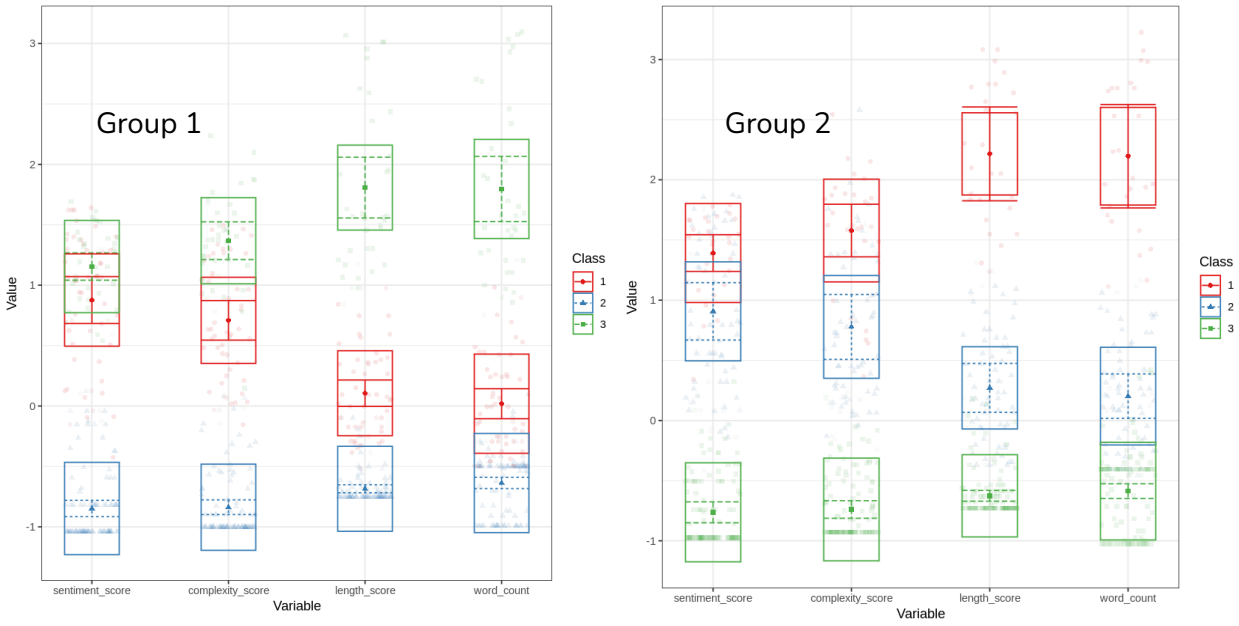
This can indicate that there could be a decrease in time invested in the module for the moderate and low engagement groups. However, a longer time series study should be conducted.



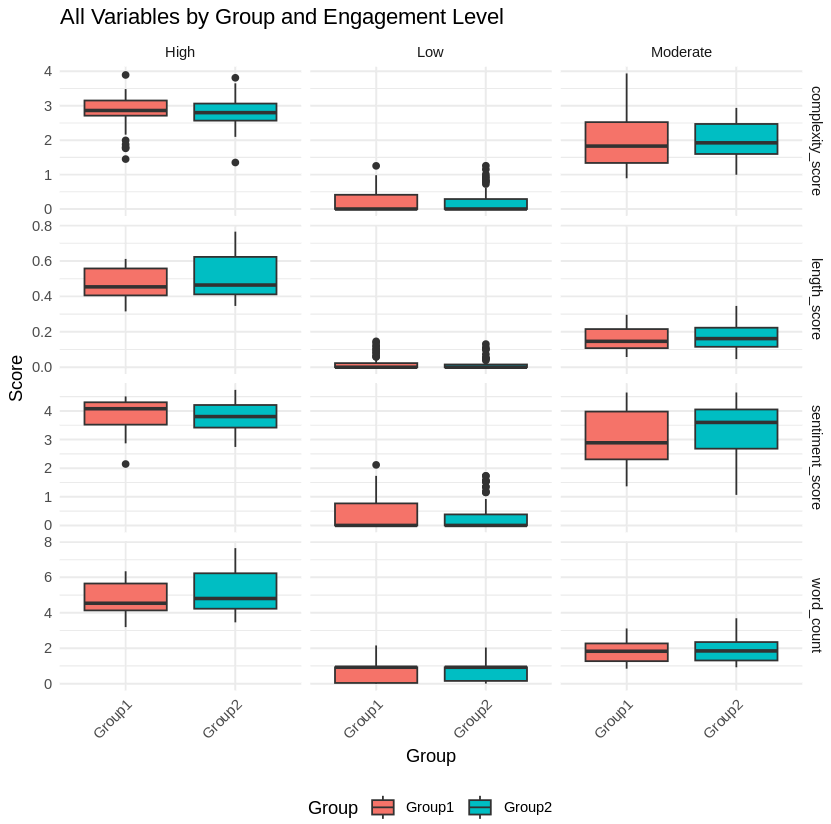
The similarity between groups in the preliminary visualization could be due to the difference in the distribution of engagement levels. Group 1 has more students in the low engagement profile whereas Group 2 has more students in the moderate profile.

Text Score Analysis

Profiles

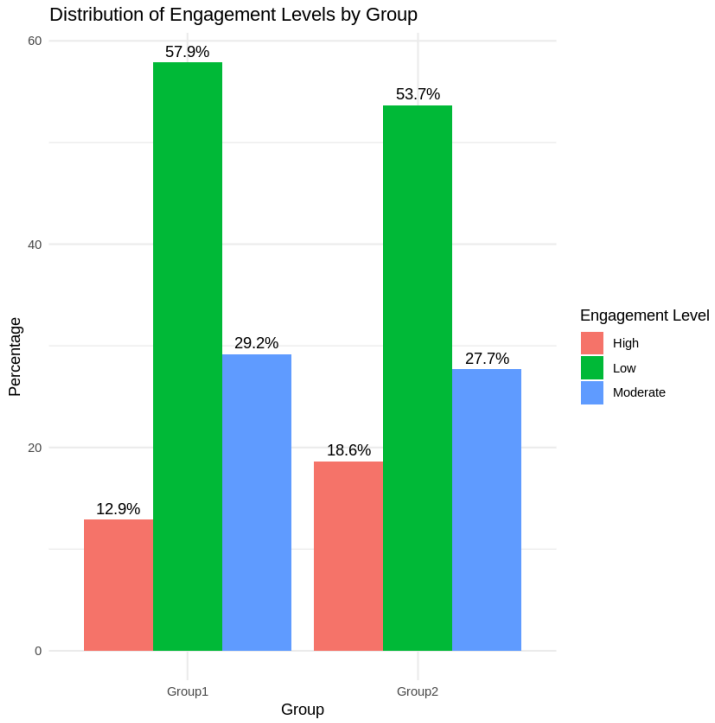


A 3 profile solution was the best fit for both groups. I then followed the same method of grouping into engagement profiles based on their distance from the mean.

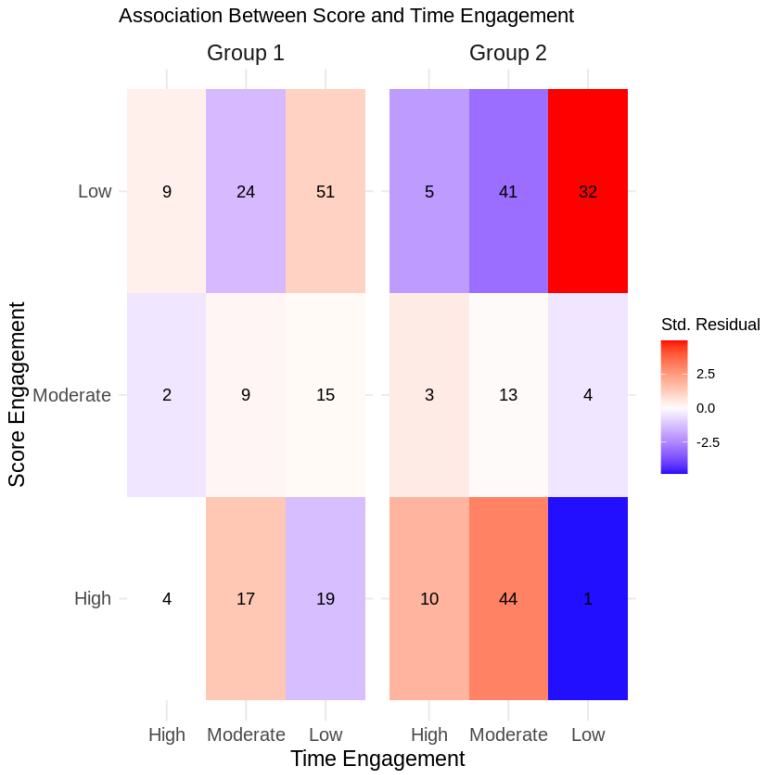


	Variable	Profile	W	p_value	effect_size	n_group1	n_group2	interpretation
W3	complexity_score	High	501.0	0.414	0.106	27	33	Small effect
W6	length_score	High	383.5	0.361	0.118	27	33	Small effect
W	sentiment_score	High	483.0	0.582	0.071	27	33	Negligible effect
W9	word_count	High	373.5	0.288	0.137	27	33	Small effect
W5	complexity_score	Low	5962.5	0.600	0.036	121	95	Negligible effect
W8	length_score	Low	6022.5	0.502	0.046	121	95	Negligible effect
W2	sentiment_score	Low	5896.5	0.716	0.025	121	95	Negligible effect
W11	word_count	Low	5673.5	0.868	0.011	121	95	Negligible effect
W4	complexity_score	Moderate	1305.0	0.256	0.108	61	49	Small effect
W7	length_score	Moderate	1348.0	0.380	0.084	61	49	Negligible effect
W1	sentiment_score	Moderate	1277.0	0.192	0.124	61	49	Small effect
W10	word_count	Moderate	1449.5	0.789	0.026	61	49	Negligible effect

For all of the profiles and variables, we have small to negligible results, with none of the results being statistically significant. Thus we cannot reject the null hypothesis as groups are similar with each other across all variables and profiles.



The similarity in groups is further emphasized through the distribution of engagement profiles, which are also consistent.



Discussion

The findings from this study present a nuanced picture of student engagement with the Mr. Kato Ophthalmology Module. Behavioural indicators—especially time spent on page types—reveal a slight shift over time toward reduced engagement among moderate and low-engagement users. While these trends are statistically significant, the effect sizes remain small to moderate, suggesting a gradual rather than abrupt decline.

Interestingly, high-engagement users in Group 2 appear more invested in Assessment-related content, which may reflect an increased emphasis on interactive components or improved instructional alignment over time. However, the small sample size of high-engagement users limits the generalizability of this result and warrants cautious interpretation.

In contrast, text score analyses indicate no meaningful change in cognitive effort across cohorts. The stability in written response quality suggests that while students may be spending less time navigating the module, their output remains consistent.

The absence of a correlation between registration duration and content engagement further emphasizes that time alone is not a reliable proxy for learning. Instead, granular metrics such as slide completion and written effort provide a richer picture of student interaction.