The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #44547

Effect of Assessment Structure on Perceived Efficacy of a Rocketry Course

Scott Nguyen
Dr. Joshua Rovey, University of Illinois Urbana-Champaign
Heather Ruth Arnett, University of Illinois Urbana-Champaign

Heather Arnett is the Coordinator of STEM Engagement Activities in Aerospace Engineering at the University of Illinois at Urbana-Champaign. She received her Master of Library and Information Science (MLIS) from the University of Illinois at Urbana-Champaign in 2018 with an emphasis on experiential STEAM learning and outreach programs. Her work focuses on recognizing resource assets and needs in diverse learning spaces and developing methods for accessible learning.

Effect of Assessment Structure on Perceived Efficacy in a Rocketry Course

Abstract

This study investigates the impact of assessment structure on student performance and engagement in an academic setting, specifically focusing on an introductory rocketry course for undergraduate non-aerospace engineering students. Departing from traditional end-of-course assessments, the research explores whether implementing a 'chunking' approach by breaking the final assessment into individual quizzes over the last week yields distinct outcomes. The approach involved comparing two groups of students: one undergoing a traditional cumulative assessment (Group A) and the other experiencing the modified ‘chunking’ assessment structure (Group B). Paired t-tests were employed to compare the results between the two groups. The results reveal that Group B outperformed Group A with a 24% increase in final assessment scores. Additionally, Group B exhibited higher levels of engagement with the material during the assessment week. These findings suggest that modifying the assessment structure by dividing the final assessment into multiple portions may reduce cognitive and testing fatigue, leading to improved student performance and increased engagement. Further research could delve into the underlying mechanisms driving these effects to inform the design of effective assessment strategies in educational settings.

Introduction

Recent studies have explored the effectiveness of distributing assessments throughout the course duration. Typically, conventional approaches involve a cumulative assessment at the course conclusion, supplemented by smaller assignments or assessments distributed at regular intervals. The primary aim of the final assessment is to evaluate students' overall knowledge acquired throughout the entire course (Cecilio-Fernandes et al., 2018; Kerdijk et al., 2015; Domenech et al., 2015; Popkova, 2018; Clark & Autar, 2021). To clarify, in the context of this study, 'cumulative assessments' refer to the practice of evaluating students' progress through continuous assessments distributed over the duration of the course.

Several investigations into the impact of cumulative assessments have yielded noteworthy results. A comparison between students subjected to cumulative assessment and those undergoing end-of-course assessment found no significant differences in overall outcomes (Cecilio-Fernandes et al., 2018). However, a distinct increase in knowledge was observed over multiple progress tests, indicating a positive trajectory in knowledge acquisition for both assessment methods (Cecilio-Fernandes et al., 2018). Notably, students exposed to cumulative assessment dedicated more time to self-study, leading to significantly improved performance on specific test items related to the later stages of the course (Kerdijk et al., 2015). This improvement persisted consistently over various weeks, suggesting that the structure of cumulative assessments influenced students' prioritization of test preparation over time (Kerdijk et al., 2015).

Frequent Cumulative Testing (FCT) is an assessment method that involves regular testing of students on the course material throughout the semester. The tests are cumulative, meaning they cover all the material taught up to that point. FCT aims to prevent last-minute cramming, promote continuous learning, and improve academic performance by providing timely feedback to students. FCT, along with continuous cumulative assessment, has been associated with improved academic performance, timely feedback, and heightened motivation among students (Domenech et al., 2015; Popkova, 2018). Notably, FCT has demonstrated superiority over traditional final exams, underscoring its positive impact on student learning and overall performance (Domenech et al., 2015). Furthermore, continuous cumulative assessment shows promise as a viable alternative to final examinations, with positive effects on both extrinsic and intrinsic student motivation (Popkova, 2018)

In a modified blended approach that incorporates evidence-based testing strategies within a numerical methods course, cumulative midterm tests preceded by practice tests demonstrated significantly higher scores for both cumulative final exams and concept inventory results (Clark & Autar, 2021).

While there is abundant research on cumulative assessments, other factors contributing to their success, such as testing fatigue due to exam length, may play a crucial role. Ploomin and Kim (2023) suggested issues with traditional end-of-course assessments leading to decreased technical knowledge post-exam due to testing fatigue. The study here proposes modifying the final assessment. Instead of a singular comprehensive exam, the final assessment is chunked over the last week by breaking it into individual quizzes. This study builds upon the cumulative assessment approach utilized by Ploomin and Kim (2023) in Group A, while incorporating a modified chunked assessment in Group B.
The primary inquiry guiding this study is whether the adoption of this modified assessment structure would produce different results. A crucial aspect of this investigation involves assessing the potential impact of revising the assessment structure to integrate best practices.

The key questions guiding this research are:

1. Does altering the final assessment structure significantly impact outcomes related to cognitive fatigue, considering variations in question timing—either in concentrated sessions or spread out over an extended period?
2. Can incorporating best practices into the assessment structure lead to distinct and potentially improved results?

Literature Review

Assessment Strategies

Education employs a wide array of assessment strategies, and recent research sheds light on their effectiveness and implications. Authentic assessment approaches, surpassing traditional methods, demonstrate a positive impact on academic achievement (Ghosh et al., 2020).
Practices aimed at preparing students for real-world applications, emphasizing collaboration and adaptability, have shown a beneficial effect (Fawns and O'Shea, 2018). Moreover, the

noteworthy acceptance of gamified assessment methods carries significant implications (Georgiou and Nikolaou, 2020).

In the realm of online education systems, the importance of monitoring students' emotions and behaviors is underscored, with effective strategies proposed, such as hidden tracking and tailored feedback (Jayasinghe et al., 2015). A notable shift from traditional exams to Paper Reviews has resulted in heightened student engagement and relevance, aligning teaching and assessment strategies with authentic research (Sletten, 2021). These studies also unveil an increased positive attitude toward online assessment methods, highlighting the necessity for purposeful planning and training to address challenges associated with new tools and scheduling conflicts (Cirit, 2015).

In summary, these collective studies emphasize the diverse and evolving landscape of assessment strategies in education, with a focus on promoting meaningful learning experiences and adapting to the changing demands of the educational environment.

Cumulative Assessments

Cumulative assessments, employed across diverse educational settings, evaluate students' overall understanding and retention of knowledge and skills acquired over time. This approach, covering material from multiple courses, facilitates a comprehensive measure of learning progression, encouraging deeper understanding (den Boer et al., 2021).

Cumulative assessments ensure minimal competency, identify knowledge gaps, and foster accountability for cumulative knowledge and skills (Vyas et al., 2015). These assessments play a crucial role in shaping comprehensive learning outcomes and aiding educators in refining instructional strategies (Muniasamy et al., 2015). Various forms of cumulative assessments, such as exams or projects, prompt students to integrate knowledge, enhancing critical thinking and problem-solving skills. Educators benefit by gaining insights into the effectiveness of teaching strategies and curriculum design (Muniasamy et al., 2015). Implementation of cumulative assessment systems, supported by information technology, enhances academic performance and training quality. These systems monitor students' progress, providing timely feedback to facilitate their learning journey (Kozlov et al., 2019).

In summary, cumulative assessments, emphasizing knowledge synthesis, play a pivotal role in shaping comprehensive learning outcomes and aiding educators in refining instructional strategies across diverse educational settings.

Cognitive Fatigue

Cognitive fatigue, marked by mental weariness, plays a pivotal role in shaping various aspects of performance (Sievertsen et al., 2016). As the day progresses, there is a noticeable decline in student test scores, underscoring the intricate link between cognitive fatigue and academic outcomes (Sievertsen et al., 2016). Importantly, the association between cognitive fatigue, negative well-being, and reduced academic achievement emphasizes the significant and independent influence of cognitive fatigue on performance, even when established predictors are taken into account (Smith, 2018). The subjective experience of fatigue intensifies with prolonged time-on-task (Ackerman and Kanfer, 2009), highlighting the importance of recognizing time constraints and the availability of cognitive resources when assessing the impact of cognitive fatigue (Borragán et al., 2017). Regardless of cognitive load, subjective cognitive fatigue increases with task duration (Sandry et al., 2014). Furthermore, the relationship between cognitive fatigue, response bias, and brain activation adds complexity to the understanding of cognitive fatigue's impact on decision-making processes (Wylie et al., 2021).

Questioning traditional assumptions regarding longer testing sessions, research suggests that incorporating additional exam items is associated with improved scores and enhanced performance, challenging the conventional beliefs about cognitive fatigue (Jensen et al., 2013). In contrast to the prevailing idea that cognitive fatigue only impairs cognitive abilities, there are instances where it may actually contribute to the facilitation of procedural motor sequence learning. This implies a more nuanced relationship between cognitive fatigue and skill acquisition (Borragán et al., 2016).

Chunking of Exams

Breaking up exams into smaller parts, known as chunking, has become a notable strategy to help students cope with challenges they may face during tests(Drexel University, n.d.). Chunking, as defined by Drexel University, involves the division of exams into sections or chunks, allowing students experiencing condition-related flare-ups during exams to complete the remaining sections within 48 hours of the original test date. This accommodation aims to maintain the integrity of exams by ensuring fairness and preventing potential advantages that may arise from accessing the entire exam at once.

Moreover, chunking finds its roots in cognitive psychology, particularly in the realm of learning and memory enhancement (University of Massachusetts Amherst, n.d.). Chunking involves breaking down complex information into smaller, more manageable parts, reducing cognitive overload and enhancing comprehension and retention(University of Massachusetts Amherst, n.d.). This is supported by Thalmann, Souza, and Oberauer (2019), who highlight how chunking in working memory tasks reduces cognitive load, thereby improving memory for other concurrent information.

Chunking serves as an effective strategy to enhance learning and memory processes (University of Massachusetts Amherst, n.d.). By organizing information into meaningful units, individuals can better process, retain, and retrieve information, ultimately improving academic performance. Furthermore, chunking reduces cognitive load, freeing up mental resources for other tasks and fostering more efficient cognitive processing (Thalmann, Souza, & Oberauer, 2019).

Background

Course Structure

In this study, we implemented a structured approach to assess student engagement and learning outcomes in technical content. Beginning with an initial evaluation of student self-efficacy and interest through surveys, we then administered a pre-course quiz to gauge baseline understanding. Following this, students engaged with the technical material, after which a mid-course quiz was conducted to evaluate learning progress. Finally, we reassessed student self-efficacy and interest after completion of all technical quizzes. This methodology provided valuable insights into the relationship between student engagement, self-efficacy, and learning outcomes, highlighting the effectiveness of periodic assessments in tracking and enhancing student progress. For details on the course schedule, please refer to **Figure *1***.

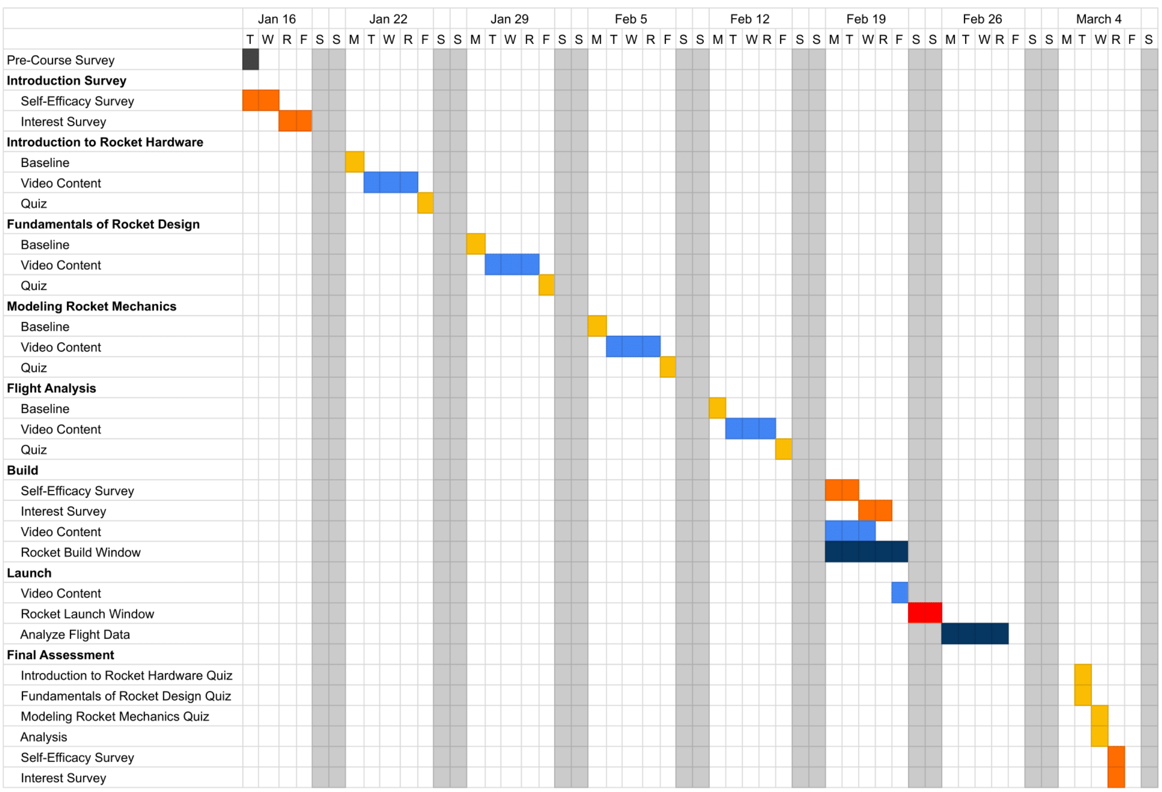


Figure 1: Course Schedule

Group A Final Exam:

Group A undergoes a cumulative final exam, characterized by a flexible timeline and sequential completion of quizzes. Over the span of a week, students are tasked with completing a series of quizzes in a predetermined order:

1. Self-Efficacy, 2. Interest, 3. Introduction and Rocket Hardware, 4. Fundamentals of Rocketry, 5. Modeling Rocket Dynamics, 6. Analysis

Importantly, students must finish each part before proceeding to the next, with no prescribed structure for when they initiate each section, provided all are completed before the final deadline.

Group B Final Exam:

In contrast, Group B experiences a "chunked" cumulative final exam, where different portions of the assessment are distributed across the week and must be completed on designated days. The final exam is divided into three distinct components:

Exam 1: Introduction and Rocket Hardware + Fundamentals of Rocketry
Exam 2: Modeling Rocket Dynamics + Analysis
Exam 3: Self-Efficacy + Interest

Students in Group B are required to complete each exam segment on its designated day, ensuring a structured approach to assessment with periodic review and reinforcement. Figure 2 describes the changes in assessment structure.

Interest and Self-Efficacy are measured using a 7-point Likert Scale, with detailed questions provided in subsequent sections of the paper. Technical quizzes are evaluated based on the percentage of correct responses. Refer to **Table *1*** for a detailed comparison of the assessment structure

Table 1: Comparison Between Assessment Structure

Initial Final Assessment Structure				
Mon	**Tue**	**Wed**	**Thur**	**Fri**
Final Exam				

Modified Final Assessment Structure				
Mon	**Tue**	**Wed**	**Thur**	**Fri**
	Exam 1).	Exam 2).	Exam 3).	

Logic Model

The Logic Model in **Table *2*** displays the connections between the exam structure and the intended outcomes (student performance and testing fatigue). It categorizes two groups: Group A, subjected to cumulative exams, and Group B, exposed to chunked exams. Cumulative exams correlate with diminished student scores and increased testing fatigue, attributed to inadequate preparation and flexible completion schedules. Conversely, chunked exams yield elevated scores and enhanced readiness, compelling students to complete exams within designated timeframes. Consequently, the logic model illustrates that the implementation of chunked exams leads to enhanced student performance and diminished testing fatigue compared to cumulative exams.

Table 2: Logic Model

Group	Inputs	Outcomes	Conclusion
Group A	- "Unchunked" Final Exam - Students finish at their own pace	- Leads to testing fatigue - Students don't prepare for exam	Lower student scores and less accurate representation of student knowledge
Group B	- "Chunked" Final Exam - Forced pace to finish on specific days	- More manageable exams - Students are able to better prepare	Higher student scores and more accurate representation of student knowledge

Theory of Change

The theory of change underlying the structuring of the final exam is designed to optimize learning outcomes and reduce testing fatigue among students in a rocketry course. By chunking the exam into smaller portions that progressively build upon foundational concepts, students are better able to digest and retain the material. This approach begins with fundamental concepts, gradually increasing in complexity through subsequent exams. The technical exams are strategically structured to create a cycle of learning, with each component building off the previous one.

Moreover, the sequencing of the exam components is intentionally designed to mitigate testing fatigue. By placing the longer and potentially more challenging sections towards the end of the exam, students are able to focus on smaller, more manageable portions initially. This approach not only enhances retention but also promotes forced spacing, allowing students to concentrate on specific topics on designated days.

Additionally, the integration of self-efficacy and interest assessment in the final exam serves to gauge students' confidence and engagement after they have completed the technical portion. This strategic sequencing ensures that students' mental bandwidth is not overly taxed at the beginning of the exam, allowing for a more accurate reflection of their attitudes and motivations towards the subject matter. Overall, this theory of change aims to optimize learning effectiveness, minimize testing fatigue, and provide a comprehensive assessment of student understanding and engagement in the rocketry course.

Methods

Researcher Positionality

Our research team is a collaborative effort that includes one master's student and a professor in aerospace engineering, and one experienced educator with expertise in outreach and educational research. This introductory rocketry course was designed specifically to spark the interest and knowledge of non-aerospace engineering freshmen and sophomores in the world of rocketry and potential space careers. The study presented here delves into a quantitative analysis of the cognitive fatigue in assessment structure.

Procedure

The research was conducted within a spring 2024 course titled "AE298: Introduction to Rocketry" offered by the Aerospace Engineering Department at a major public university in the United States. Recruitment involved various advertising across the university, including departmental emails, strategically placed flyers, and outreach to 4 undergraduate engineering student groups. The target audience was first- and second-year STEM students outside of Aerospace Engineering. Course participation, encompassing quizzes and surveys, students received two credit hours (all participants in this study received full credit). However, these credits were insufficient to fulfill technical elective requirements, typically demanding three or more credits. The interview questions were strategically categorized into two groups: self-efficacy and interest. Self-efficacy questions aimed to gauge students' confidence in their ability to master the course material and apply their learning. Interest questions, on the other hand, explored their engagement with the topics, personal connections they formed, and intrinsic motivation to delve deeper. This two-pronged approach aimed to uncover nuanced insights into student experiences, potentially revealing hidden gaps or strengths that surveys might miss.

Data Analysis

In data analysis, selecting the appropriate statistical test is crucial for accurate conclusions. When data deviates from normality, a rank sum test is preferred, while a parametric t-test is suitable for normally distributed data. Both aim to derive a p-value, typically with a significance level (alpha) set at 0.01 for a 99% confidence level, minimizing Type I errors.

Interpreting results follows a clear guideline: if the p-value is below alpha, the null hypothesis is rejected, suggesting a significant difference. Conversely, a p-value above alpha retains the null hypothesis, implying no significant difference. In hypothesis testing, the null suggests Group B outperforms Group A, with the alternative suggesting no significant difference.

For assessing changes over time, a paired t-test is ideal, estimating mean differences with a null hypothesis of zero change. Rejection of this null, indicated by a p-value below alpha, signifies a significant change.

To address cognitive and testing fatigue, diverse methods are employed, including performance metrics, response time analysis, error rates, and qualitative surveys. These aim to identify trends indicating fatigue and compare data across different assessment structures to pinpoint significant differences.

Results

Student Demographics

Details and demographics of the student groups utilized in the two studies are outlined in **Table *3*** for comparison. Group A is comprised of thirty-two undergraduate students, with 63% being male and 34% female. The majority were either Asian (59%) or white (38%), with 59% being first-year students. Fields of study included mechanical engineering (28%) and physics (25%), the latter being part of the engineering program. Additionally, three students from outside engineering—two from mathematics and one from business—were included. The selection criteria favored early-stage college participants, resulting in 87.5% being freshmen and sophomores, and 12.5% being juniors and seniors.

Group B is comprised of twenty-six undergraduate students was examined. This group consisted of 58% male and 42% female students, with the majority being either Asian (62%) or white (38%). Furthermore, 58% were first-year students, and the fields of study included mechanical engineering (31%) and astrophysics (15%), with physics being part of the engineering curriculum. Similar to the first study, early-stage college participants were favored in selection, resulting in mostly freshmen and sophomores (85%), with a smaller proportion being juniors and seniors (15%).

Table 3: Group A and B Student Demographics

	Group A		Group B	
Categories	n	Percentage	n	Percentag e
Total	32	100	26	100
Gender				
Female	11	34.40	11	42.31
Male	20	62.50	15	57.69
Prefer not to say	1	3.10	0	0.00
Ethnicity				
Do not wish to provide	1	3.12	0	0.00
Hispanic or Latino/a	6	18.80	3	11.54
Not Hispanic or Latino/a	25	78.10	23	88.46
Race (Multiple selections allowed)				
American Inidian or Alaska Native Asian	1	3.12	0	0
Asian	19	59.38	10	38.46
Black or African American	2	62.50	0	0
White	12	37.50	16	61.54
Do not wish to provide	1	3.12	0	0
Year in College				
1	19	59.40	15	57.69
2	9	28.10	7	26.92
3	3	9.40	3	11.54
4	1	3.12	1	3.85
College				
Agricultural and Biomedical Engineering	1	3.12		
Astrophysics			4	15.38
Astronomy			3	11.54
Civil Engineering	1	3.12		
Chemical Engineering			2	7.69
Computer Engineering			1	3.85
Computer Science	1	3.12		
Electrical and Computer Engineering	2	6.24	2	7.69
Engineering Mechanics			1	3.85
Engineering Undeclared	1	3.12		
Industrial and Enterprise Systems	4	12.50	1	3.85
Engineering	2	6.24		
Material Science and Engineering	9	28.10	8	30.77
Mechanical Engineering	2	6.20	1	3.85
Math			1	3.85
Nuclear Engineering	8	25.00	2	7.69
Physics	1	3.12		
Business				

Interest

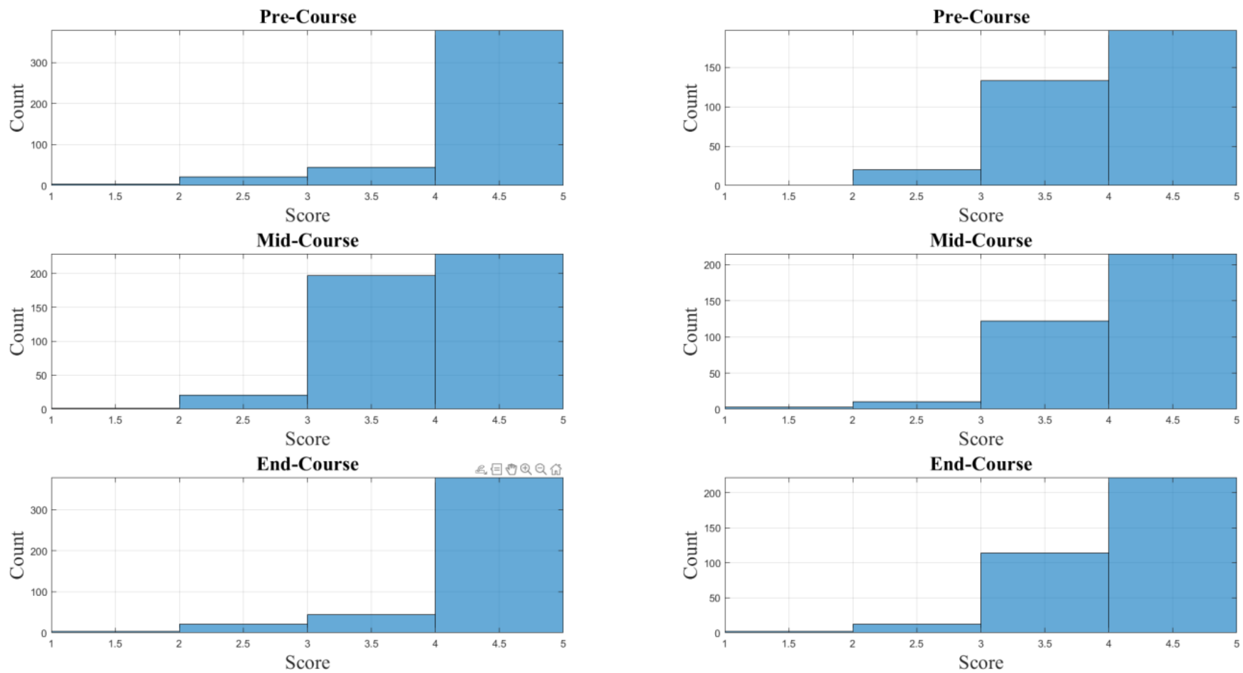


Figure 2: **Group A(left) and Group B(right) Interest Survey HistogramFigure** *2* displays histograms representing the score distributions from the Interest Survey analysis for both participant groups. Notably, both groups consistently demonstrate high levels of interest throughout the course duration. The distribution of scores deviates from a normal pattern, instead favoring a left-skewed distribution.

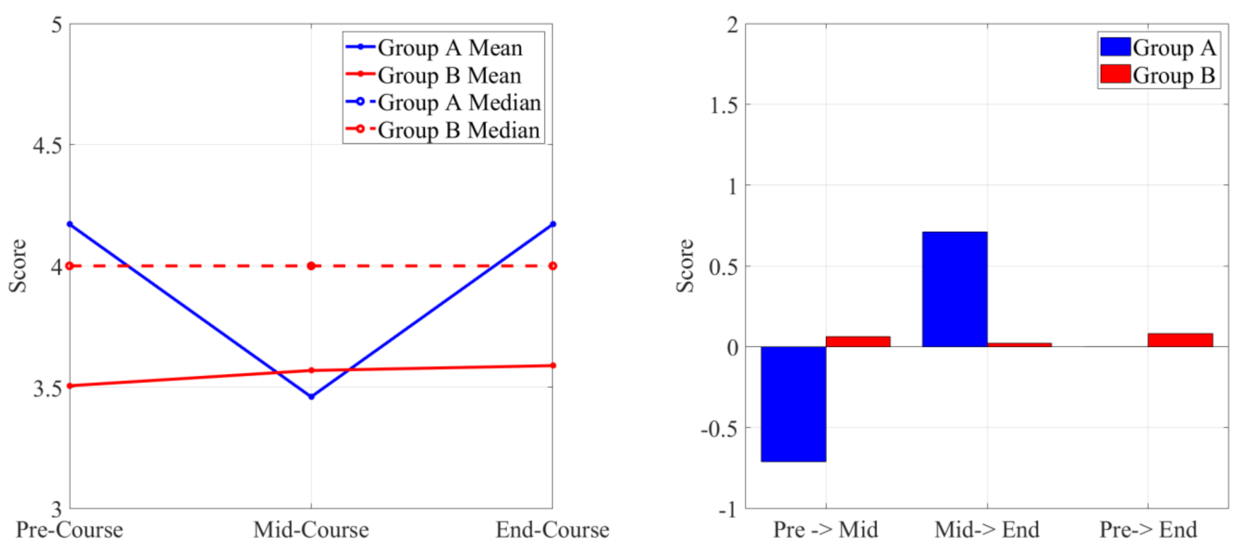


Figure 3: Interest Survey Mean and Median Comparison(left) Change in Mean(right)

Figure *3* illustrates the mean and median scores over the course duration for both groups, along with the change in mean scores across course sections. While both groups maintained consistent median interest levels throughout the course, indicating no median interest change, Group A exhibited a decline in mean interest during the middle phase, rebounding to initial levels by the course end. Conversely, Group B demonstrated stable mean interest levels, mirroring their median scores. However, Group B showed slight incremental mean interest increases over the course duration.

Table 4: Hypothesis Results for Interest Levels between Groups

	P-Value	**Reject Null**
Pre-Course	1.810e-5	Yes
Mid-Course	7.316e-1	No
End-Course	2.894e-5	Yes

Table 5: Mean and Median Interest Levels between Groups

	Mean			**Median**		
	Pre	**Mid**	**End**	**Pre**	**Mid**	**End**
Group A	4.17	3.46	4.17	4.00	4.00	4.00
Group B	3.51	3.57	3.59	4.00	4.00	4.00

Table *4* presents the hypothesis test results for interest level comparisons between groups. The null hypothesis is that Group A exhibits higher interest levels, while the alternative suggests no difference in interest between groups. Statistical analysis indicates that at the beginning and end of the course, the "no Chunked" Group A indeed demonstrates higher interest levels compared to Group B. However, no significant evidence supports this difference during the middle phase of the course.

Table 6: Group A Hypothesis Results for Change in Interest Levels

	P-Value	**Reject Null**
Pre to Mid	9.073e-8	Yes
Mid to End	8.074e-8	Yes
Pre to End	5.026e-8	No

Table 7: Group B Hypothesis Results for Change in Interest Levels

	P-Value	Reject Null
Pre to Mid	3.702e-1	No
Mid to End	3.771e-1	No
Pre to End	1.955e-1	No

Self-Efficacy

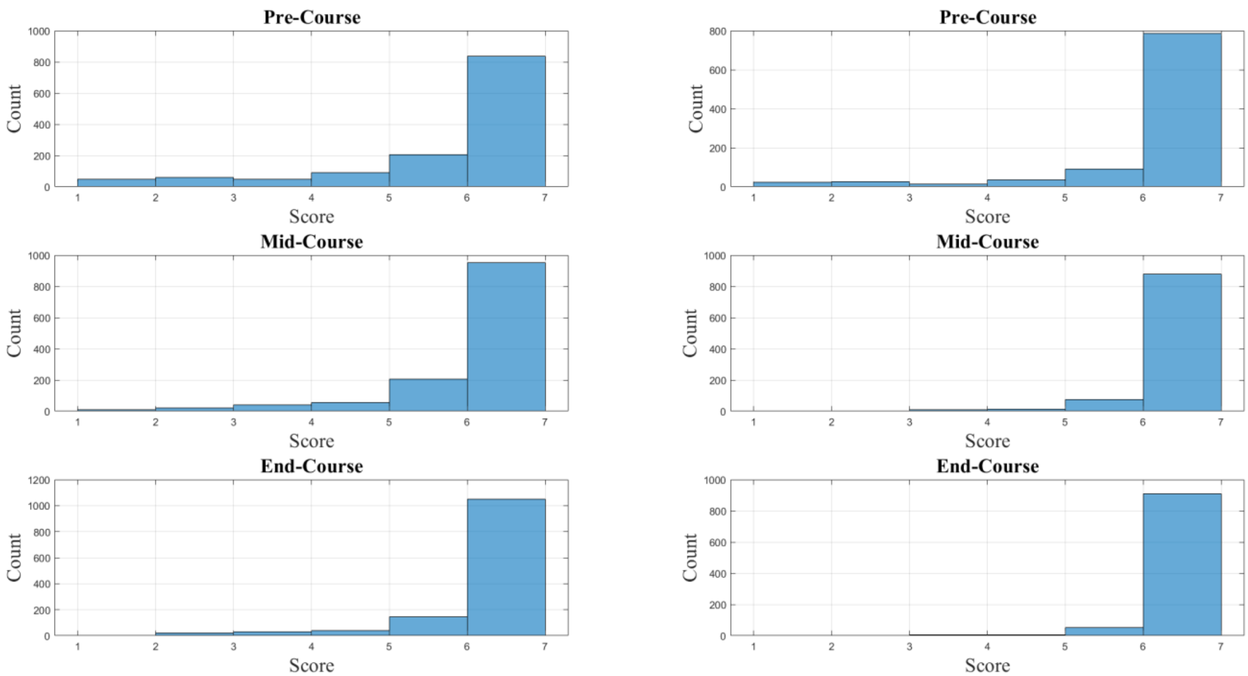


Figure 4: Group A(left) Group B(right) Self-Efficacy Histogram

When comparing self-efficacy levels, a similar pattern emerges as seen in the interest surveys. Both groups show high levels of self-efficacy. It's clear from the histograms that self-efficacy continues to rise, with more students reporting higher levels of self-efficacy.

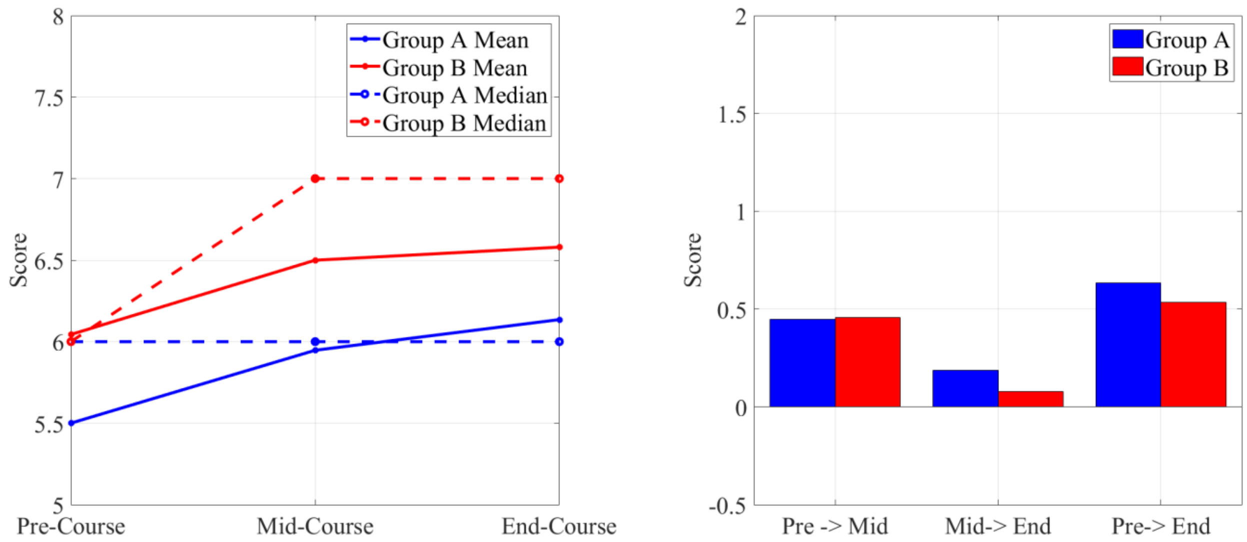


Figure 5: Mean and Median Comparison (left) Change in Mean (right) for Self-Efficacy Survey

Throughout the entire course, it's clear that self-efficacy steadily rises for both groups. This stands in contrast to Group A's interest levels, which decreased midway through the course, while Group B's interest remained relatively stable.

Table 8: Hypothesis Results for Self-Efficacy Levels

	P-Value	**Reject Null**
Pre-Course	5.905e-3	Yes
Mid-Course	2.122e-2	No
End-Course	4.644e-3	Yes

Table *8* presents the results of hypothesis testing for comparing self-efficacy levels among different groups. The null hypothesis suggests that there is no difference in self-efficacy levels between Group A and Group B, while the alternative hypothesis proposes that Group B exhibits higher self-efficacy levels. Statistical analysis indicates that at the beginning and end of the course, Group B indeed demonstrates higher self-efficacy levels compared to Group A. However, there is no significant evidence supporting this difference during the middle phase of the course.

Table 9: Group A Hypothesis Results for Change in Self-Efficacy Levels

	P-Value	**Reject Null**
Pre to Mid	1.369e-3	Yes
Mid to End	1.349e-1	No
Pre to End	9.901e-5	Yes

Table 10: Group B Hypothesis Results for Change in Self-Efficacy Levels

	P-Value	**Reject Null**
Pre to Mid	2.147e-3	Yes
Mid to End	2.017e-1	No
Pre to End	1.005e-5	Yes

Technical Quiz Histograms

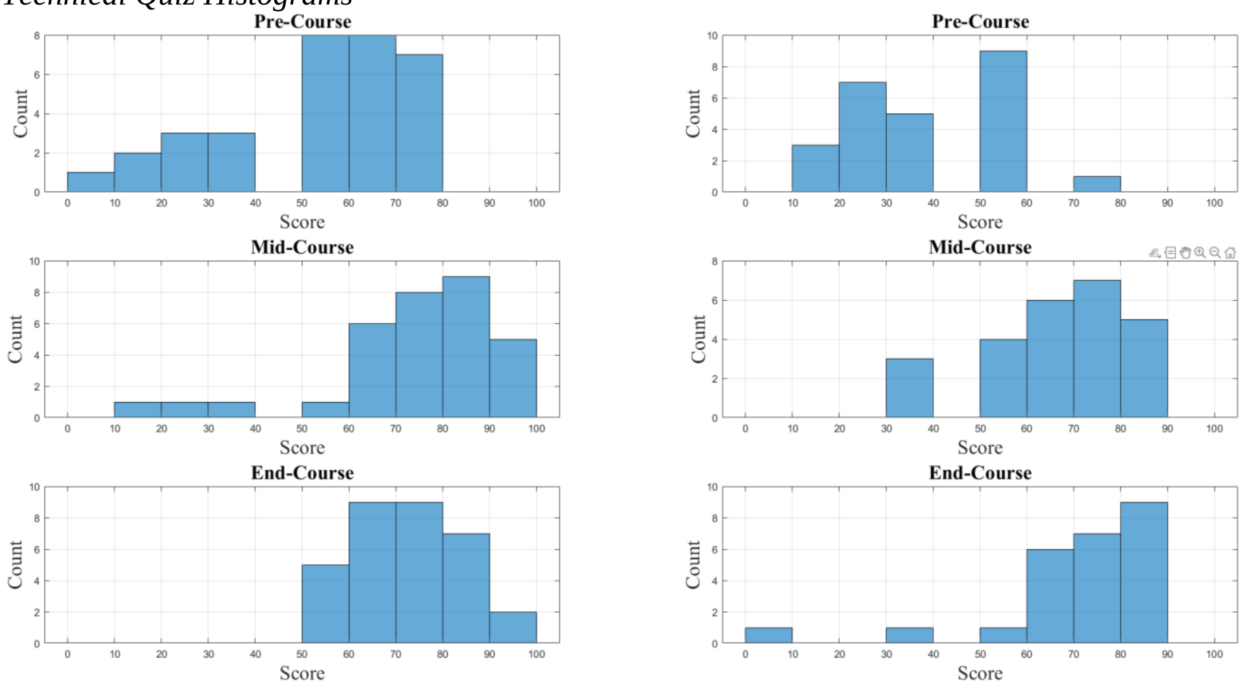


Figure 6: Group A(left) Group B(right) Introduction and Rocket Hardware Quiz Histogram

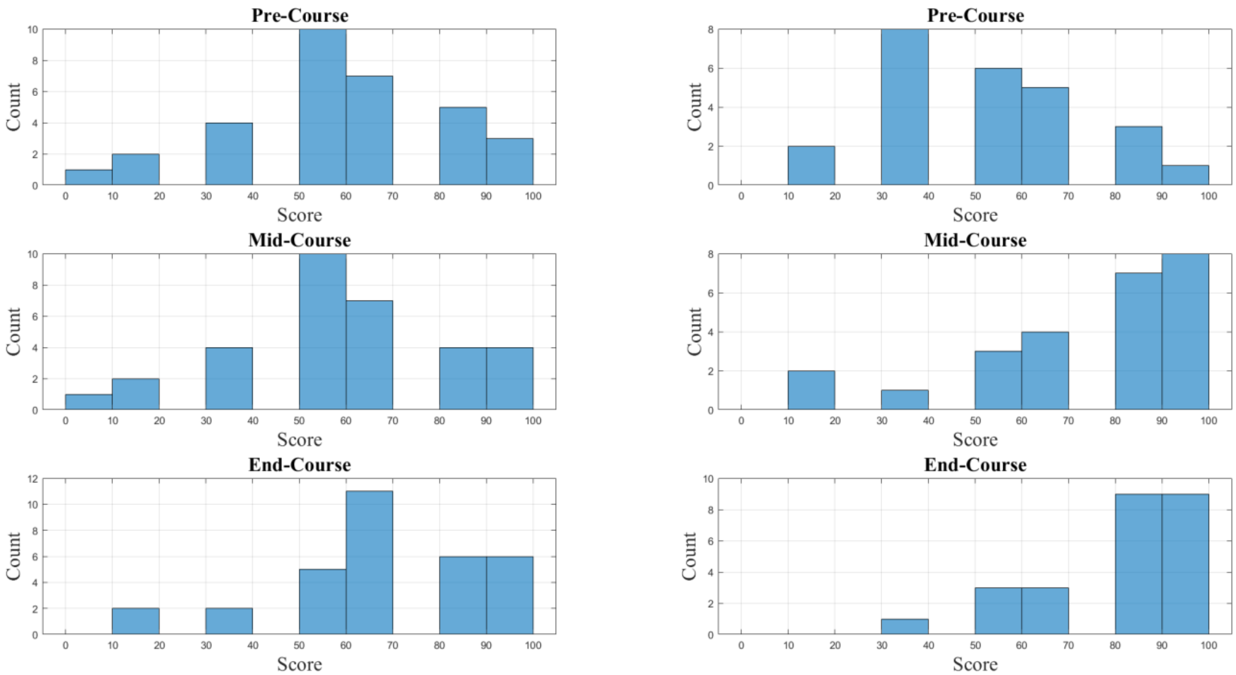


Figure 7: Group A(left) Group B(right) Fundamentals of Rocketry Quiz Histogram

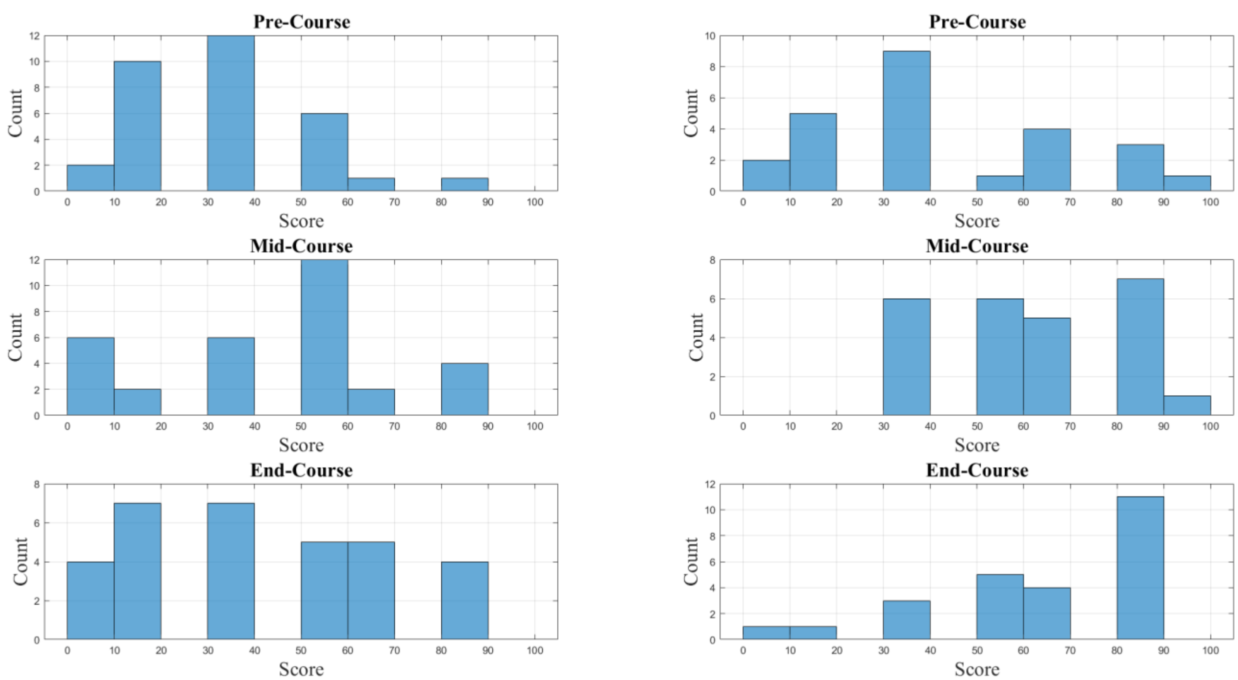


Figure 8: Group A(left) Group B(right) Modeling Rocket Mechanics Quiz Histogram

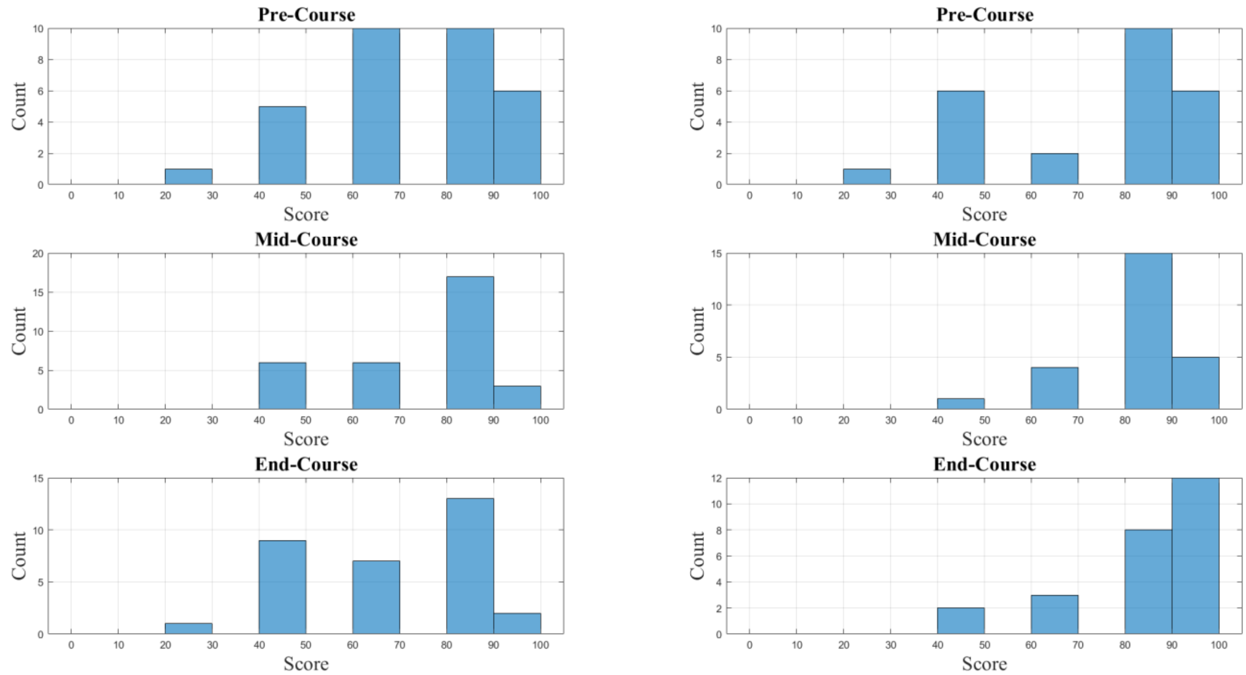


Figure 9: Group A(left) Group B(right) Analysis Quiz Histogram

Figure *6*-*Figure* 9shows the histogram of scores for during the Pre-Course, Mid-Course, and End-Course assessments for each module. At a glance we can generally see scores starting off low at for the pre-course assessment and improving during the Mid-Course assessment.

Technical Quiz Mean and Change in Mean Plots

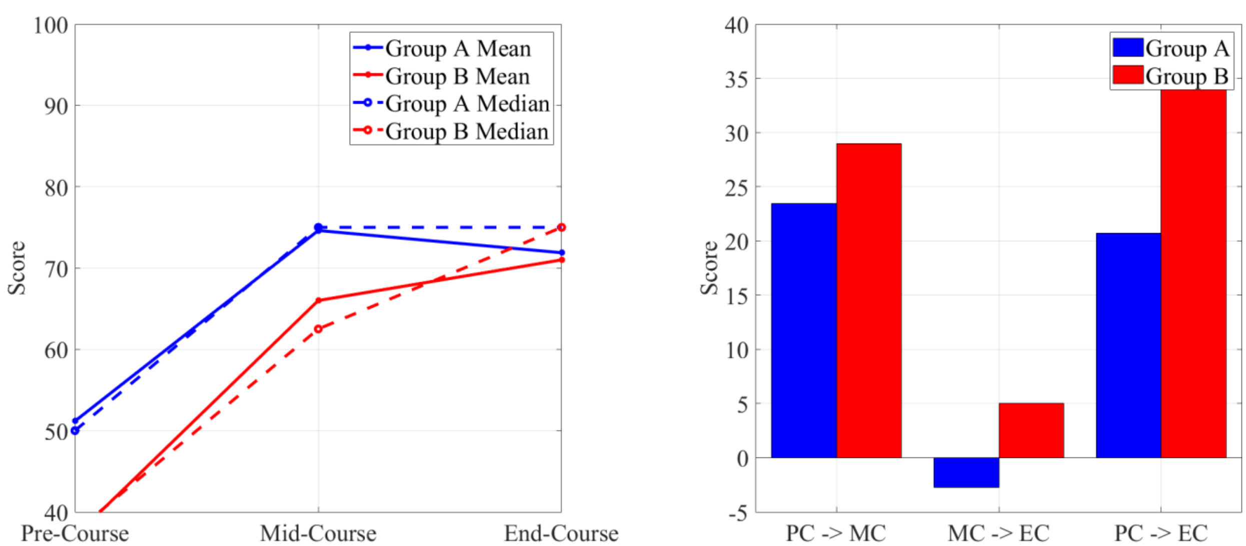


Figure 10: Introduction and Rocket Hardware Mean and Median Comparison (left) Change in Mean (right)

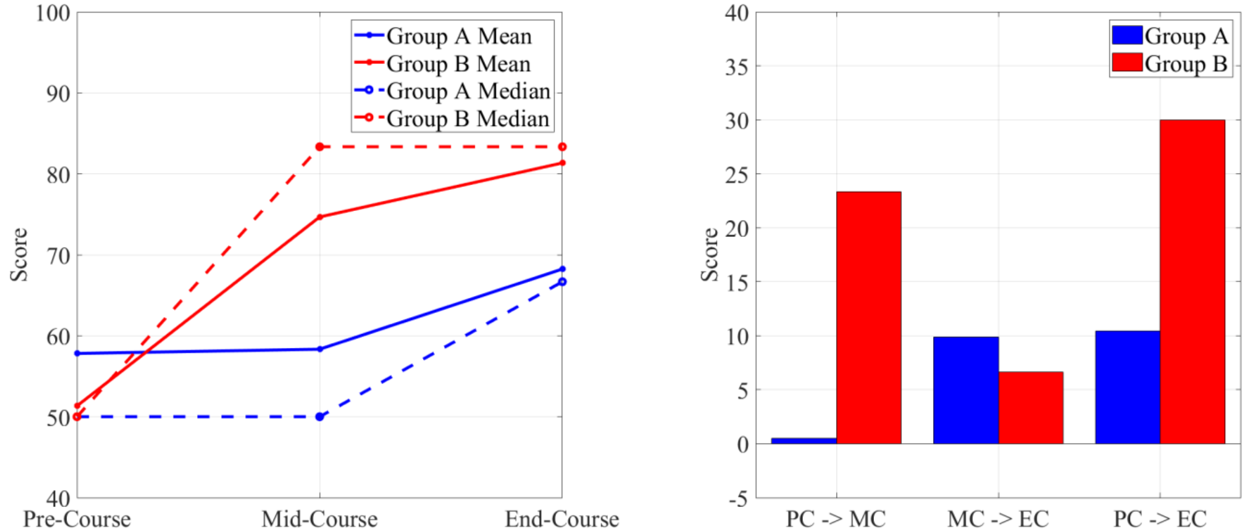


Figure 11: Rocketry Fundamentals Mean and Median Comparison (left) Change in Mean (right)

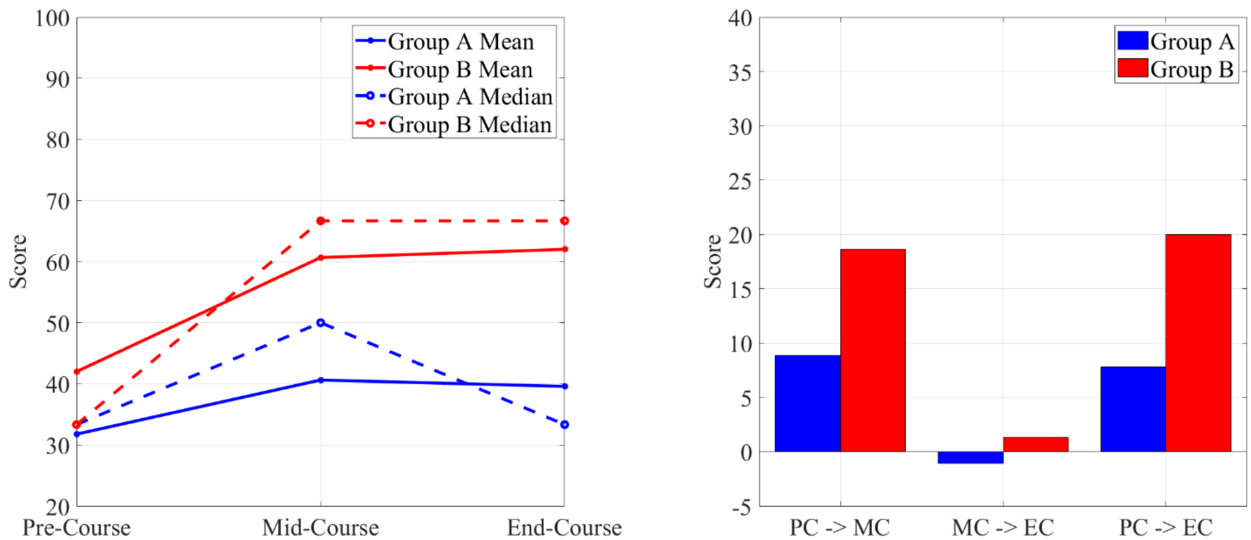
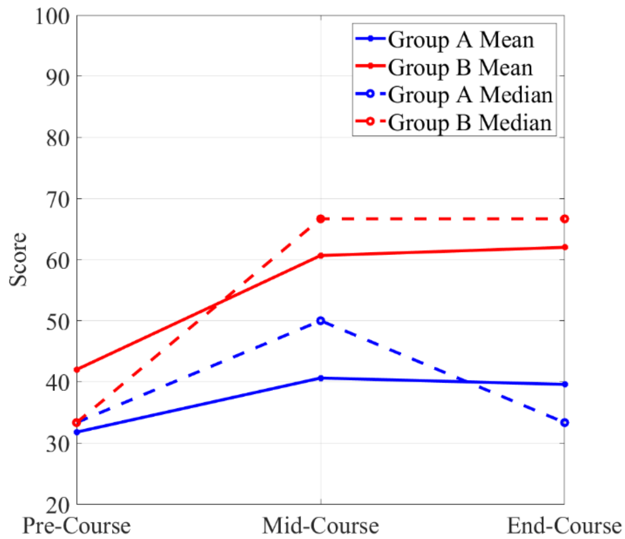


Figure 12: Modeling Rocket Dynamics Mean and Median Comparison (left) Change in Mean (right)



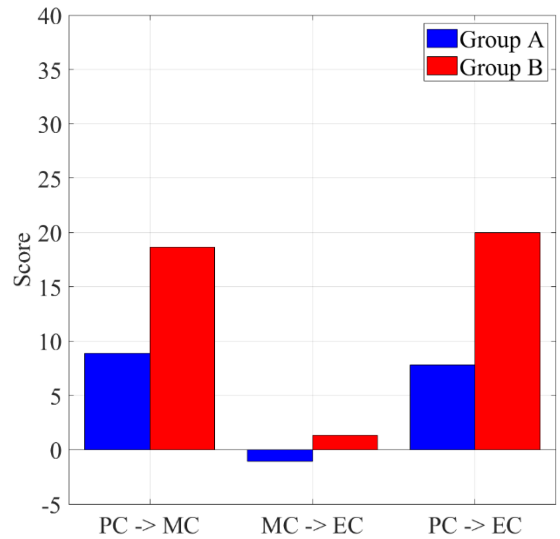


Figure 13: Analysis Mean and Median Comparison (left) Change in Mean (right)

Figure *10-Figure 13*show the mean and median for each group as well as the change in mean scores. We can see that there is an increase in scores during from Pre-Course to Mid-Course which is expected after students view the online content. When comparing Mid-Course to End-Course values, we see that scores for Group A decrease while scores for Group B increase. Tabulated scores are presented in the following section.

Technical Quiz Mean and Change in Mean Tables

Table 11: Rocketry Fundamentals Mean and Change in Mean

	Mean			Change in Mean		
	Pre	Mid	End	Pre to Mid	Mid to End	Pre to End
Group A	51.17	74.61	71.88	23.44	-2.73	20.70
Group B	37.00	66.00	71.00	29.00	5.00	34.00

Table 12: Rocketry Fundamentals Mean and Change in Mean

	Median			Change in Median		
	Pre	Mid	Pre	Pre to Mid	Mid to End	Pre to End
Group A	57.81	58.33	68.23	0.52	9.90	10.42
Group B	51.33	74.67	81.33	23.33	6.67	30.00

Table 13: Modeling Rocket Dynamics Mean and Change in Mean

	Mean			Change in Mean		
	Pre	Mid	End	Pre to Mid	Mid to End	Pre to End
Group A	31.77	40.63	39.58	8.85	-1.04	7.81
Group B	42.00	60.67	62.00	18.67	1.33	20.00

Table 14: Analysis Mean and Change in Mean

	Median			Change in Median		
	Pre	Mid	End	Pre to Mid	Mid to End	Pre to End
Group A	69.38	70.62	63.75	1.25	-6.88	-5.62
Group B	71.20	79.20	84.00	8.00	4.80	12.80

Technical Quiz Hypothesis Test

Table 15: Introduction and Rocket Hardware Hypothesis Results

	P-Value	**Reject Null**
Pre-Course	4.516e-1	No
Mid-Course	7.295e-1	No

Table 16: Rocketry Fundamentals Hypothesis Results

	P-Value	**Reject Null**
Pre-Course	8.862e-1	No
Mid-Course	6.655e-1	No

Table 17: Modeling Rocket Dynamics Hypothesis Results

	P-Value	**Reject Null**
Pre-Course	8.402e-1	No
Mid-Course	9.123e-1	No

Table 18: Analysis Hypothesis Results

	P-Value	**Reject Null**
Pre-Course	8.402e-1	No
Mid-Course	9.123e-1	No

Table *15*-*Table 18* show the p-value for the hypothesis test. The Null is that there is no difference in mean scores between the groups and the alternative being there is a difference. In every case, the results show that there is no evidence to suggest that the two groups performed differently for during the Pre-Course and Mid-Course assessments.

Final Exam Results

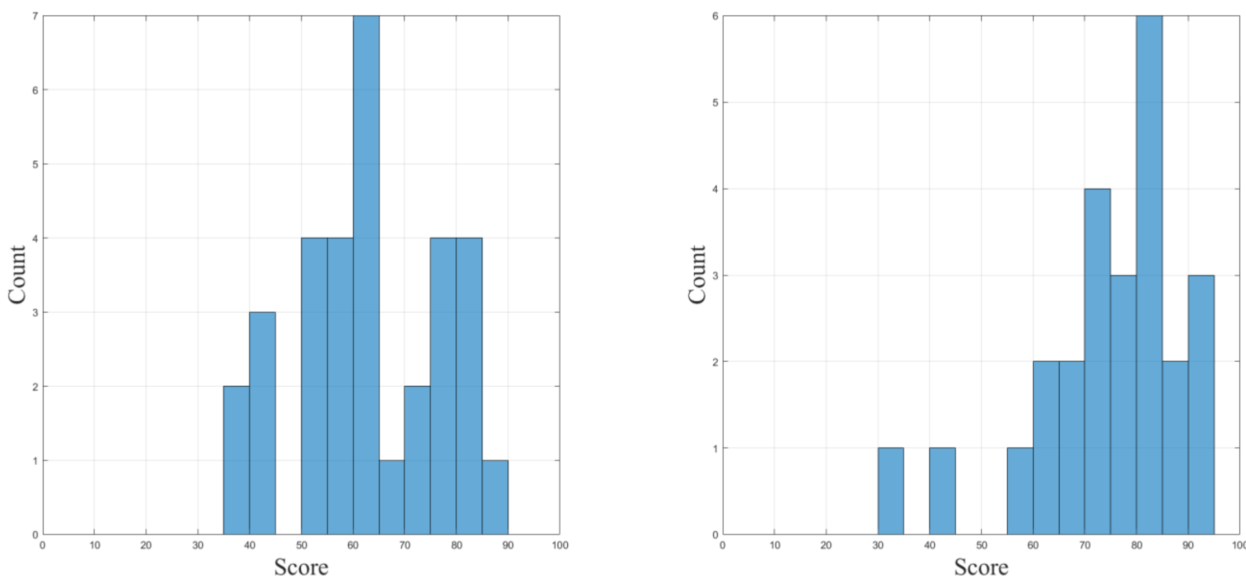


Figure 14: Group A(left) and Group B(right) Final Exam Score Histogram

Table 19: Final Exam Mean and Median

	Mean			Change in Mean		
	Pre	**Mid**	**End**	**Pre to Mid**	**Mid to End**	**Pre to End**
Group A	69.38	70.62	63.75	1.25	-6.88	-5.62
Group B	71.20	79.20	84.00	8.00	4.80	12.80

Table 20: Final Exam Hypothesis Results

	P-Value	Reject Null
Final Exam	1.056e-13	Yes

In **Figure *14***we can see that the Group A’s final exam is approximately normal with a mean of 61.62 while Group B has a left skewed distribution with a mean of 73.92 and a median of 76. For this hypothesis test the null states that Group B has higher final exam scores than Group A and the alternative is that Group B doesn’t have higher final exam scores than Group B. The results from the hypothesis test show that there is significant evidence to suggest that Group B has higher final exam scores than Group A.

Student Course Engagement During Course

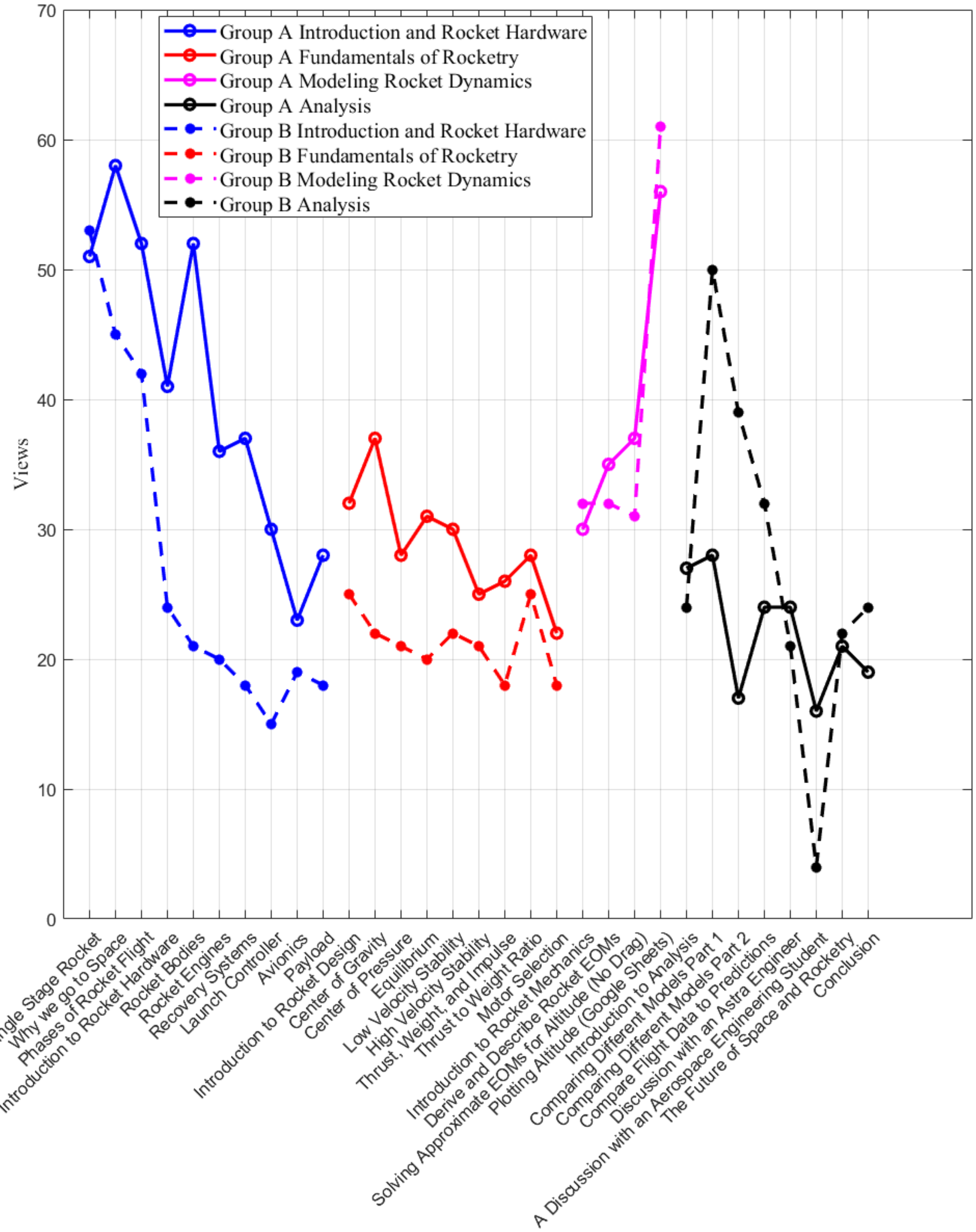


Figure 15: Views per Video Comparison during Courses

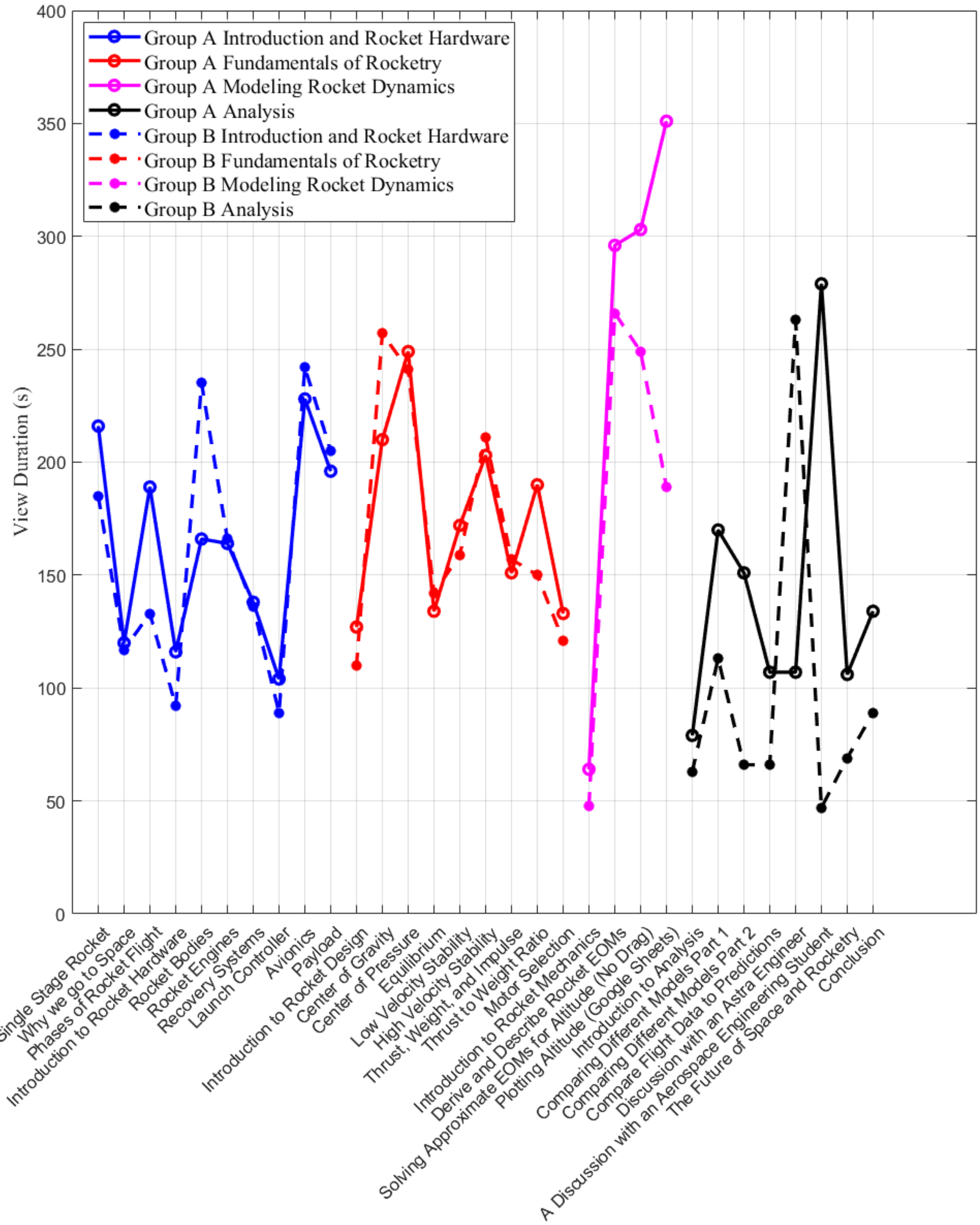


Figure 16: View Duration per Video Comparison during Course

Figure *15* and **Figure *16***show the total amount of views per video and average view duration per video respectively for the first four weeks of the course. As shown in Figure 1., each module aligns with a certain week. The data was collected for the week that the module was assigned as well as the week prior to encompass any students that may have started viewing the content earlier. There is a large spike for the “Plotting Altitude (Google Sheets)”.

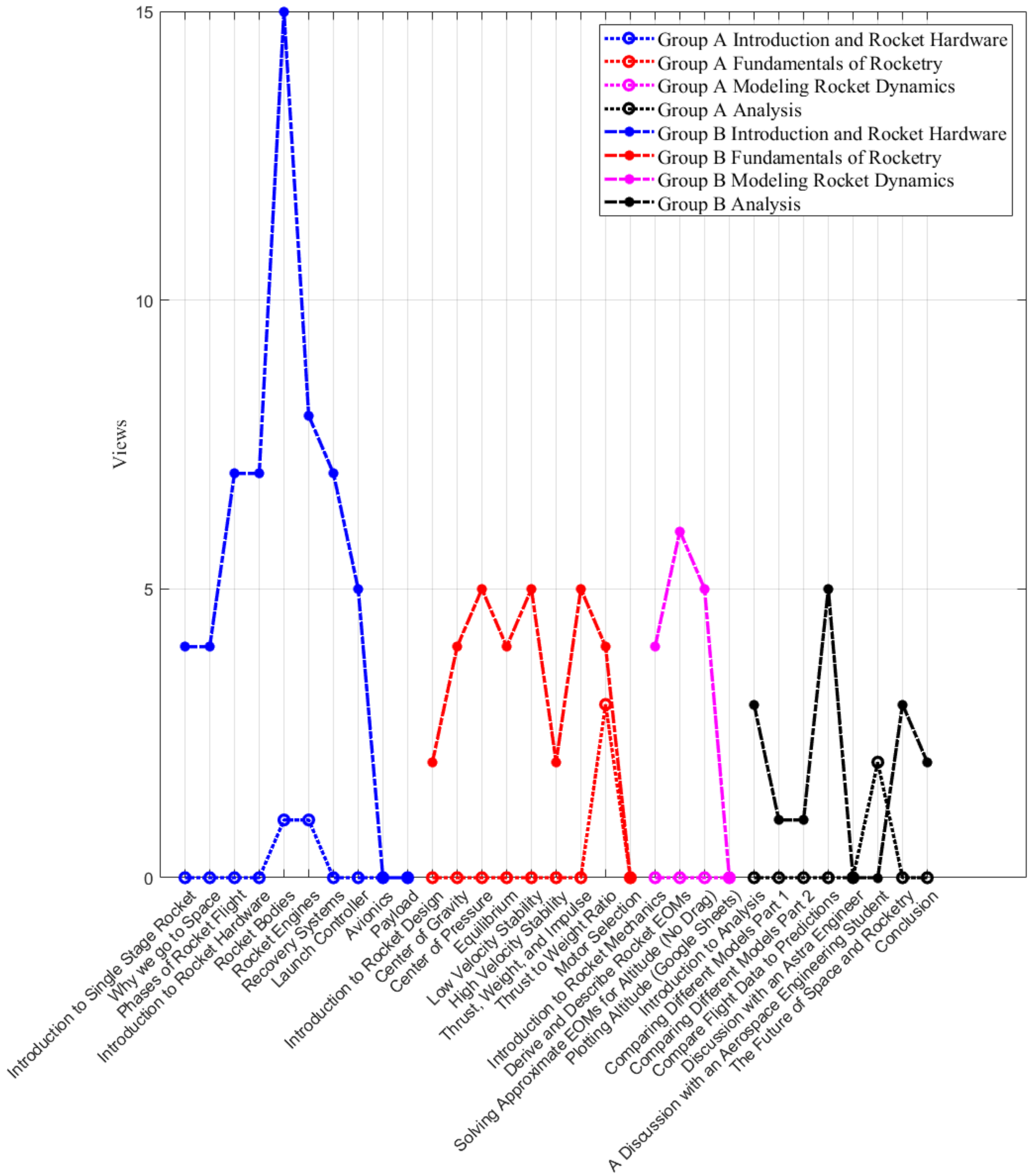


Figure 17: Views per Video Comparison during Finals Week

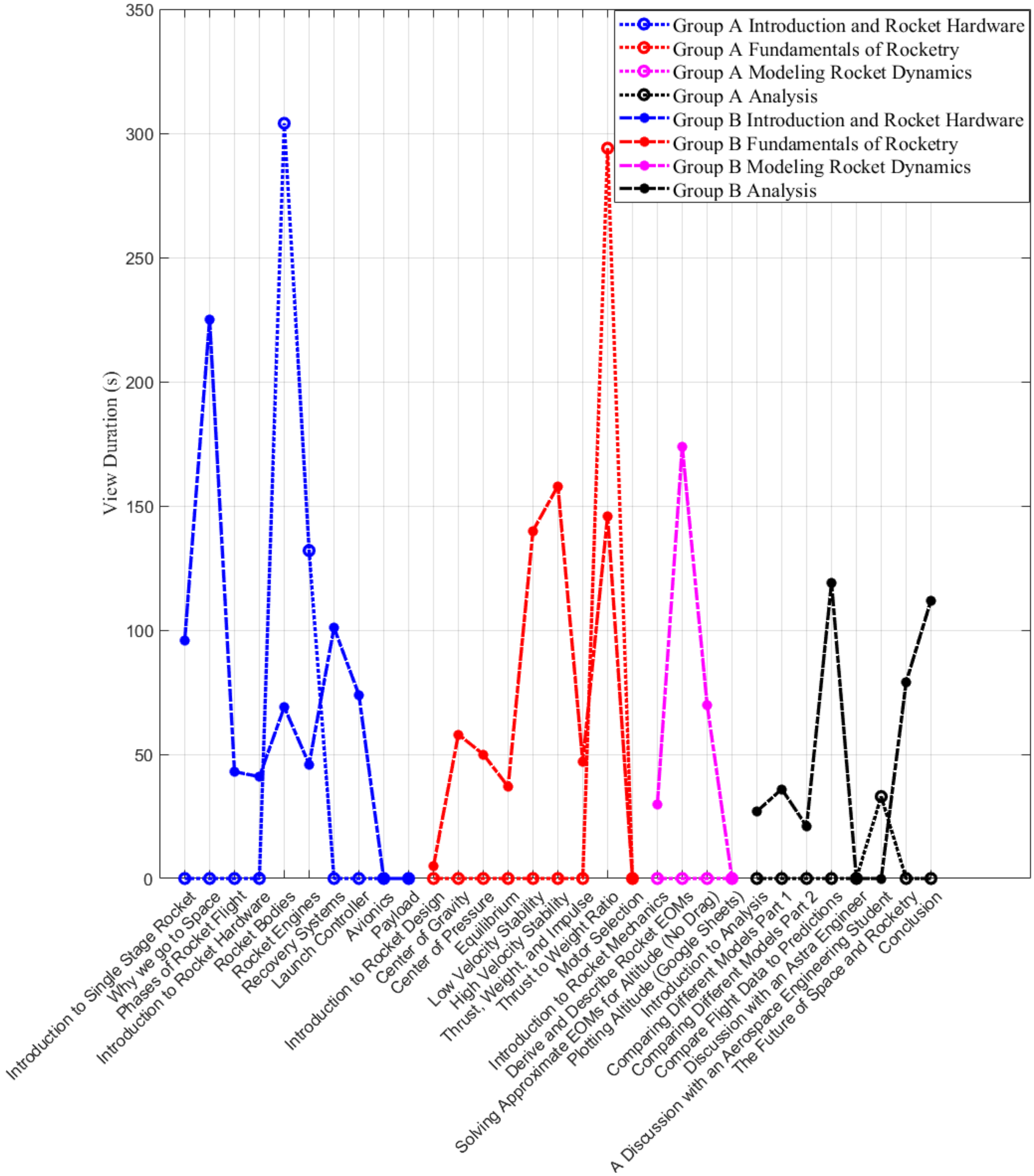


Figure 18: View Duration per Video during Finals Week

Figure *18* and **Figure 19** now show the number of views and view duration for each group during finals week. In contrast to **Figure *15*** and **Figure *16***, there is a clear difference between both groups. Group A rarely engaged with the course material compared to Group B.

Table 21: View and View Duration Hypothesis Results

	P-Value	**Reject Null**
Course Views Amount	8.402e-1	No
Course View Duration	9.123e-1	No
Finals Views Week Amount	9.086e-09	Yes
Finals Week View Duration	1.4720e-06	Yes

Table *21* shows the hypothesis test results comparing the view amount and view duration. For the course, the null is that there is no difference between the view amount and view duration between both groups, with the alternative being there is a difference. The p-value is greater than our alpha value indicating that there is no evidence to suggest that there is a difference in the view amount and view duration during the course. During finals week, the null is that Group B is more engaged with the course content and the alternative is that there is no difference. With a p-value much less than alpha, there is evidence to suggest that Group B was more engaged with the course content during finals week than Group A.

Finals Week Completion Rate

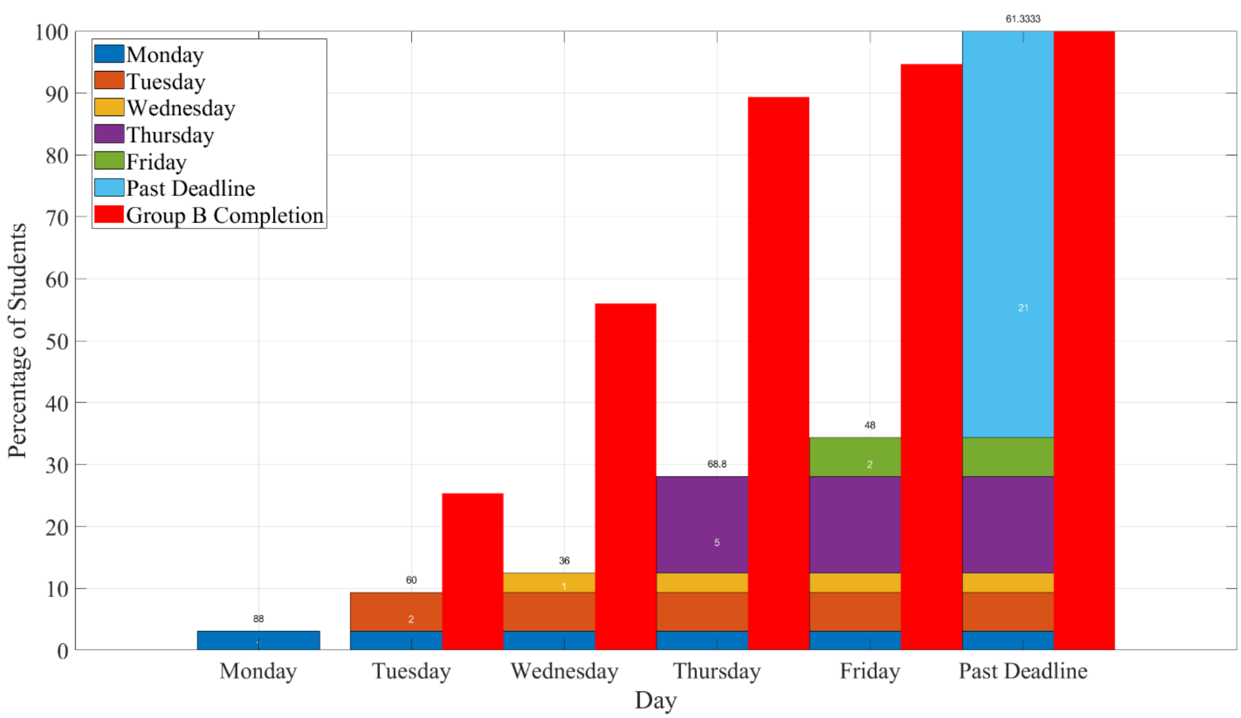


Figure 19: Finals Week Completion Rate

Table 22: Group A Finals Week Completion Rate and Mean Score

Day	Number of Students	Mean Score
Monday	1	88
Tuesday	2	60
Wednesday	1	36
Thursday	5	69
Friday	2	48
Past Deadline	21	61

Figure 19 and **Table 22** display the completion rate of the final exam between both groups. Group B students for the most part followed the prescribed chunked exam structure with only a few students submitting after the deadline. Over 50% of students from Group A finished the final exam after the deadline.

Discussion

Our main objective in our data analysis is to note any key differences between the two groups prior to the final exam and any differences in the groups during the final exam. This is done to ensure that both groups are similar so that we can attribute any findings to the modified final assessment structure.

Comparison of Interest and Self-efficacy

Table 3 presents a notable disparity in interest levels between Group A and Group B throughout the Pre and End Course periods, consistently showing Group A with higher interest levels. Conversely, Table 4 indicates a different trend, with Group B exhibiting higher levels of self-efficacy. Existing literature underscores the reciprocal relationship between interest and self-efficacy (Niemivirta & Tapola, 2007), where stronger self-efficacy tends to coincide with heightened interest, suggesting that enhancements in self-efficacy can positively impact task performance by bolstering interest (Niemivirta & Tapola, 2007). Moreover, while self-efficacy directly shapes individuals' goal-setting behaviors, interest may indirectly enhance self-regulation when coupled with self-efficacy (Lee, Lee, & Bong, 2014). Hence, both interest and self-efficacy are deemed vital components for academic success.

However, our findings diverge from the literature. Group A exhibited a positive shift in interest levels from Pre to Mid Course and Mid to End Course, albeit with a decline during the Mid-Course phase. Notably, Group A consistently displayed higher interest levels during the Pre and End Course phases, in contrast to Group B, which exhibited negligible fluctuations in interest levels throughout the course duration.

In line with existing literature, these variations should be mirrored in the self-efficacy outcomes, suggesting that Group A should attain higher scores during the Pre and End Course phases due to their elevated interest levels. Interestingly, technical quiz scores showed no discernible differences between the two groups.

Turning to self-efficacy, Group B showcased higher levels during the Pre and End Course phases, with both groups experiencing similar increases from Pre to Mid and Pre to End Course phases, and minimal changes during the Mid to End Course phase. Considering these self-efficacy results, we would anticipate a parallel pattern in interest levels. Additionally, since Group B demonstrated higher self-efficacy levels during the Pre and End Course phases, one would expect them to outperform Group A during those stages. Notably, only the final assessment during the End Course phase revealed notable score discrepancies between the two groups.

The observation that there were no significant score differences between the groups until the final assessment leads us to two conclusions. First, both self-efficacy and interest equally impact student performance. Second, the variance in final assessment scores is attributed solely to changes in the assessment structure rather than fluctuations in self-efficacy or interest, or a combination of these factors.

Interest and Self-Efficacy Survey

Group A's interest levels at the beginning and end of the course may correspond with Rottinghaus, Larson, and Borgen's (2003) findings of a moderate relationship between self-efficacy and interest. Their research suggests that self-efficacy influences interest development through mastery experiences, with a threshold effect where moderate self-efficacy is necessary to sustain interest. This implies that Group A's initial interest could stem from their perceived ability to succeed. Group B's consistently higher self-efficacy may contribute to their sustained engagement and confidence throughout the course.

Technical Quizzes Scores

The hypothesis tests conducted on the Pre-Course and Mid-Course quizzes indicate no significant difference in scores between the two groups. This suggests that students from both groups commence the course with equivalent levels of knowledge, and upon engaging with the course material, they conclude with comparable levels of understanding as well.

Final Exam Scores

The results of the final exam hypothesis suggest that Group B outperformed Group A. Additionally, the technical quiz hypothesis test revealed no disparity in scores from the beginning of the course to after students engaged with the online content and were assessed, indicating similar performance levels between both groups with comparable demographics. The distinguishing factor contributing to Group B's higher final exam scores appears to be the assessment structure, leading us to conclude that this factor likely influenced the observed outcome discrepancy between the two groups.

Finals Week Completion Rate

The modified assessment structure significantly influenced how students completed and performed on the final exam. In Group A, where students could choose when to finish the exam, many delayed until after the deadline. Additionally, Group A's final exam required students to complete lengthy Interest and Self-Efficacy Surveys before tackling technical content, likely causing mental fatigue. This, combined with the requirement to finish earlier assessments before moving on, added to the challenge and may have affected students' performance accuracy, especially since most completed the exam late.

On the other hand, Group B had a structured exam with set deadlines. They also took a different approach, by instead starting with the initial two technical quizzes on the first day (13 questions each), followed by the final two technical quizzes on the second day (11 questions each), and concluding with the Self-Efficacy and Interest Surveys on the last day (39 questions and 14 questions, respectively).This change aimed to reduce fatigue and improve performance accuracy. Placing the Interest and Self-Efficacy Surveys at the end of the exam allowed students to reflect on their learning experience, potentially enhancing preparation and performance. Hypothesis testing revealed a significant difference in performance, with Group B outperforming Group A in the final exam. This difference was attributed to the structured exam format, which reduced fatigue and allowed for more effective preparation through manageable study segments. Additionally, the structured format enabled focused review on specific topics, enhancing students' understanding of the course material.

Conclusion

In conclusion, our examination of the two study groups revealed notable distinctions in their scores, particularly regarding their final exam scores and completion rates. Despite comparable levels of knowledge and engagement throughout the course, Group B outperformed Group A in the final assessment. This discrepancy can be primarily attributed to the modified assessment structure, which proved influential in shaping students' exam preparation and scores.

The Interest and Self-Efficacy Surveys highlighted distinct motivational and confidence dynamics between the groups, with Group A's initial interest potentially driven by perceived self-efficacy, while Group B demonstrated sustained engagement and confidence levels. However, the technical quiz scores indicated no significant initial differences between the groups, suggesting equivalent levels of baseline knowledge and engagement.

Importantly, the modified final assessment structure significantly influenced students' completion rates and exam preparation strategies. Group A, with flexible exam completion deadlines and a sequential exam structure, faced challenges related to procrastination and potential testing fatigue. In contrast, Group B's structured exam format and strategic approach to exam preparation, which involved starting with technical quizzes, and concluding with the Self-Efficacy and Interest Survey, and reviewing specific topics during finals week, contributed to improved scores.

The study sheds light on how changes in assessment structure can impact student performance. It offers insights into the relationship between assessment structure in terms of chunking exams and also the order in which students are tested. However, the study has limitations. It's just one instance in a specific course and focuses on STEM undergraduates. Replicating it across various courses and with non-STEM undergraduates would strengthen its findings. In terms of repeatability, the study provides a clear methodology that other researchers can follow. Conducting replication studies with larger samples and across different institutions would enhance the understanding of these relationships in diverse contexts, contributing to educational assessment literature.

For future research, several questions arise:

1. Do students who performed poorly on the launch revisit the material for study purposes, while those who performed well do not?
2. Is there a correlation between launch performance and subsequent video watch behavior?
3. Does the effectiveness of students' experience with the launch correlate with their final scores?
4. Does providing regular performance feedback on grades relate to improved future performance?
5. How do reminders or encouragement to complete quizzes on time impact scores and overall performance?

Addressing these questions could offer valuable insights into optimizing assessment strategies and enhancing student learning outcomes. Further exploration of these areas could inform the development of more effective assessment practices and learning support for students.

References

1. Ploomin T, Kim J. The Impact of a Blended Course about Rocketry on Student Knowledge, Self-Efficacy, and Career Interest

2. Cecilio-Fernandes D, Nagtegaal M, Noordzij G, Tio RA. Cumulative assessment: Does it improve students' knowledge acquisition and retention? Sci Med. 2018;28(4):ID31880. DOI: 10.15448/1980-6108.2018.4.31880Cecilio-Fernandes D, Nagtegaal M, Noordzij G, Tio RA. Cumulative assessment: Does it improve students' knowledge acquisition and retention? Sci Med. 2018;28(4):ID31880. DOI: 10.15448/1980-6108.2018.4.31880

3. Kerdijk, W., Cohen-Schotanus, J., Mulder, B. F., Muntinghe, F. L., & Tio, R. A. (2015). Cumulative versus end-of-course assessment: effects on self-study time and test performance. Medical education, 49(7), 709-716.

4. Domenech, J., Blazquez, D., De la Poza, E., & Muñ oz-Miquel, A. (2015). Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain. Educational Assessment, Evaluation and Accountability, 27, 153-169.

5. Tio, R. A., Stegmann, M. E., Koerts, J., van Os, T. W., & Cohen-Schotanus, J. (2016). Weak self-directed learning skills hamper performance in cumulative assessment. Medical Teacher, 38(4), 421-423.

6. Popkova, E. (2018). Continuous cumulative assessment in higher education: Coming to grips with test enhanced learning. Revisiting the Assessment of Second Language Abilities: From Theory to Practice, 331-349.

7. Clark, R. M., & Autar, K. A. W. (2021). Enhancing Student Outcomes in a Blended Numerical Methods Course for Engineers: The Case for Practice and Cumulative Tests. International Journal of Engineering Education, 37(3), 585-93.

8. Ghosh, S., Brooks, B., Ranmuthugala, D., & Bowles, M. (2020). Authentic versus traditional assessment: an empirical study investigating the difference in seafarer students' academic achievement. The Journal of Navigation, 73(4), 797-812.

9. Fawns, T., & O'Shea, C. (2018, May). Distributed learning and isolated testing: tensions in traditional assessment practices. In Proceedings of the 2018 Networked Learning Conference (pp. 132-139).

10. Georgiou, K., & Nikolaou, I. (2020). Are applicants in favor of traditional or gamified assessment methods? Exploring applicant reactions towards a gamified selection method. Computers in Human Behavior, 109, 106356.

11. Jayasinghe, U., Dharmaratne, A., & Atukorale, A. (2015, February). Students' performance evaluation in online education system Vs traditional education system. In Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation (REV) (pp. 131-135). IEEE.

12. Sletten, S. R. (2021). Rethinking assessment: Replacing traditional exams with paper reviews. Journal of Microbiology & Biology Education, 22(2), 10-1128.

13. Cirit, N. C. (2015). Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment. Turkish Online Journal of Educational Technology-TOJET, 14(3), 9-19.

14. Vyas, D., Halilovic, J., Kim, M. K., Ravnan, M. C., Rogan, E. L., & Galal, S. M. (2015). Use of cumulative assessments in US schools and colleges of pharmacy. Pharmacy, 3(2), 27-38.

15. den Boer, A. W., Verkoeijen, P. P., & Heijltjes, A. E. (2021). Comparing formative and summative cumulative assessment: Two field experiments in an applied university engineering course. Psychology Learning & Teaching, 20(1), 128-143.

16. Muniasamy, V., Ejalani, I. M., & Anandhavalli, M. (2015). Predicting the Students Learning Outcome Based on Comparing the Assessment Methods in DiplomaE-Commerce Course, Community College, King Khalid University, Ksa. Assessment, 7(7.18), 1-595.

17. Kozlov, V., Alontseva, E., & Guryanov, A. (2019, May). INFORMATION AND EDUCATIONAL ENVIRONMENT FOR MONITORING AND FORMING A CUMULATIVE ASSESSMENT OF A STUDENT’S ACADEMIC WORK IN A SEMESTER. In SOCIETY. INTEGRATION. EDUCATION. Proceedings of the

International Scientific Conference (Vol. 5, pp. 383-392).

18. Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students’ performance on standardized tests. Proceedings of the National Academy of Sciences, 113(10), 2621-2624.

19. Rashtchi, M., Abbaszadeh, M., & Tabarraee, N. (2015). Reading Comprehension Testing Techniques and Cognitive Fatigue: A Study on Construct Irrelevant Factors. Br. J. Educ. Soc. Behav. Sci, 6, 196-209.

20. Borragán, G., Slama, H., Bartolomei, M., & Peigneux, P. (2017). Cognitive fatigue: A time-based resource-sharing account. Cortex, 89, 71-84.

21. Borragán, G., Slama, H., Destrebecqz, A., & Peigneux, P. (2016). Cognitive fatigue facilitates procedural sequence learning. Frontiers in human neuroscience, 10, 86.

22. Smith, A. P. (2018). Cognitive fatigue and the well-being and academic attainment of university students. Journal of Education, Society and Behavioral Science.

23. Beck, D. (2023). Virtual school students and state testing: A story of fatigue, irrelevant thoughts and anxiety. Available at SSRN 4665005.

24. Sandry, J., Genova, H. M., Dobryakova, E., DeLuca, J., & Wylie, G. (2014). Subjective cognitive fatigue in multiple sclerosis depends on task length. Frontiers in neurology, 5, 214.

25. Wylie, G. R., Yao, B., Sandry, J., & DeLuca, J. (2021). Cognitive Fatigue and Brain Activation: A Study of the Relationship Between Cognitive Fatigue and Response Bias. Frontiers in Psychology, 11, Article 579188.

26. Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. Journal of Experimental Psychology: Applied, 15(2), 163.

27. Jensen, J. L., Berry, D. A., & Kummer, T. A. (2013). Investigating the effects of exam length on performance and cognitive fatigue. PloS one, 8(8), e70270.

28. Rottinghaus, P. J., Larson, L. M., & Borgen, F. H. (2003). The relation of self-efficacy and interests: A meta-analysis of 60 samples. Journal of Vocational Behavior, 62(2), 221-236.

29. Niemivirta, M., & Tapola, A. (2007). Self-efficacy, interest, and task performance: Within-task changes, mutual relationships, and predictive effects. Zeitschrift für Pädagogische Psychologie, 21(3/4), 241-250.

30. Lee, W., Lee, M. J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. Contemporary Educational Psychology, 39(2), 86-99.

31. Chunking of Exams. [Drexel University]. Retrieved from https://drexel.edu/disability-resources/~/media/Drexel/Treasurer-Group/Disability-Resources/Documents/Chunking-of-Exams.ashx

32. University of Massachusetts Amherst. (n.d.). How do I chunk content to increase learning? [University of Massachusetts Amherst Center for Teaching and Learning]. Retrieved from

33. Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45(1), 37.

34.