EFFECT OF ASSESSMENT STRUCTURE ON PERCEIVED EFFICACY IN A ROCKETRY COURSE

BY

SCOTT TRONG NGUYEN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Aerospace Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Adviser:

Professor Joshua L. Rovey

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, a well-known and researched educational strategy that supports more inclusive evaluation, yields distinct outcomes by breaking the final assessment into individual quizzes over the last week.

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. Both groups had similar demographics and scores prior to the assessment. The results reveal that Group B outperformed Group A with on average 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.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Defense Education Program (NDEP) for Science, Technology, Engineering, and Mathematics (STEM) Education, Outreach, and Workforce Initiative Programs under Grant No. HQ00342010040. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Department of Defense nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

I extend my deepest gratitude to my advisor, Dr. Joshua L. Rovey, for the invaluable opportunity to collaborate with him on the DoD NDEP project. Dr. Rovey's exceptional mentorship, insightful guidance, and unwavering support have been fundamental to my academic and professional development throughout this journey. His expertise and encouragement have profoundly shaped my understanding of this field and enhanced my research skills. I am also sincerely thankful to Heather Arnett for her significant contributions to the development and execution of this project, as well as her invaluable assistance in preparing my paper for future publications.

TABLE OF CONTENTS

1 INTRODUCTION

The pursuit of effective assessment strategies in education is an ongoing process. Educators and researchers are constantly in search of methods that not only accurately gauge student knowledge but also promote learning and retention. Traditional end-of-course assessments, long-standing fixtures in academic evaluation, are being re-examined in the context of our evolving understanding of cognitive processes. Research into test length and cognitive fatigue revealed that while students may feel more fatigued the longer, they work on tasks, this does not necessarily lead to a decline in performance, suggesting that individual differences and self-regulation strategies play a significant role in educational outcomes [1].

Day long break during exams can significantly improve grades indicating the beneficial effects of rest on cognitive performance [2]. Despite subjective feelings of fatigue, performance did not decrease even after extensive testing periods, pointing to the resilience of cognitive function and the influence of individual traits on fatigue and affective responses [3]. These studies collectively underscore the need to rethink traditional assessment methods in favor of strategies that consider the complex dynamics of cognitive fatigue and individual differences.

Chunking is an educational strategy that enhances the learning process by breaking down complex information into smaller, digestible segments, which is especially beneficial in online learning environments. In the absence of a traditional classroom setting, chunking has been shown to maintain student engagement and attention. Students preferred chunk-style videos, which led to higher engagement and better completion rates, and even appeared to improve assessment outcomes [4]. Chunking helps students process and understand material more efficiently by dividing content into manageable parts [5]. Chunking course materials improved exam pass rates, outperforming other methods such as expanded homework or automated feedback [6]. These studies suggest that chunking not only facilitates learning but also has a direct positive impact on student success.

The effectiveness of cumulative assessment in education is complex, with studies highlighting both benefits and challenges. This has been shown to have a positive impact on student engagement, performance, and study habits. Frequent Cumulative Testing (FCT) improves academic outcomes and pass rates [7]. While increased self-study time and better retention of recent material [8]. However, others have found no significant differences in knowledge acquisition between cumulative and end-of-course assessments, despite benefits from repetitive testing and spaced learning [9]. Researchers have also cautioned that cumulative assessments may disadvantage students with weaker self-directed learning skills, suggesting the need for additional support within such frameworks [10].

This study investigates the impact of assessment structure on student performance and engagement in an introductory rocketry course designed for undergraduate non-aerospace engineering students. Central to this investigation is whether an assessment approach based on the principles of chunking leads to different and potentially better outcomes. The key questions guiding this research are:

1. Does 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?

2 LITERATURE REVIEW

2.1 ASSESSMENT STRATEGIES

Educational assessment strategies are a key area of interest in contemporary research, particularly in relation to capturing student performance and engagement. Authentic assessments, which simulate practical scenarios, significantly enhance academic achievement, particularly for students with relevant work experience [11]. Researchers have also argued assessments that foster adaptability and critical thinking are essential in our rapidly changing world [12]. Prior studies have also compared two types of assessment strategies. Summative Cumulative Assessment (SCA)—where grades contribute to final exam scores and emphasize performance for certification—and Formative Cumulative Assessment (FCA), which focuses on feedback for learning improvement without affecting final grades. While SCA showed a short-term improvement in exam grades, its long-term benefits were inconsistent, suggesting that FCA's approach to continuous learning might offer more sustainable academic development [13]. These studies collectively advocate for a shift from traditional to more dynamic and authentic assessment methods to better prepare students for future challenges.

Engaging assessment methods can have a positive effect. Assessments integrated into engaging online games have been observed to elicit favorable reactions from participants [14]. This preference for engaging methods is supported by research showing that students found learning more relevant and were more engaged when traditional exams were replaced with paper reviews in a microbiology course [15]. Building on this idea, incorporating such methods into virtual learning environments has demonstrated significant benefits, including increased student engagement and improved examination outcomes, further emphasizing the value of innovative assessment strategies [16]. Similarly, rethinking assessment strategies to incorporate collaborative approaches resulted in significant improvements in student experiences and module grades, highlighting the benefits of innovative assessment methods [17]. These findings suggest that innovative and continuous assessment strategies are crucial for fostering student interaction, confidence, and academic success.

Personalized feedback and alternative assessment methods are widely regarded as beneficial for students, yet their effectiveness depends on thoughtful implementation and integration into the course structure. Research has explored their effects, advocating for strategies that can be tailored to student needs to maximize learning outcomes. For instance, monitoring students' emotional states to provide meaningful and tailored feedback could be especially valuable in an online course where face-to-face interactions are limited [18]. However, this requires careful planning and the incorporation of mechanisms for continuous feedback, which may be absent in the current course structure. Similarly, alternative assessments have been shown to foster positive attitudes, particularly in contexts where thoughtful planning and training are used to implement online assessments [19]. In our case, if feedback were more continuous and personalized, it could have contributed significantly to enhancing student engagement and performance [20].

Moreover, students in online courses have been found to prefer teacher feedback over peer feedback, with adaptive feedback systems shown to boost learning outcomes. Implementing structured feedback loops, which may currently be lacking in the course, could help students feel more supported and improve their learning experiences. The absence of discussion boards or collaboration elements in the course may also limit opportunities for peer interaction, which could have contributed to a more positive attitude toward the assessments and the content. Reflecting on these studies, it seems clear that integrating continuous, personalized feedback and exploring alternative assessments could significantly enhance the course and its learning outcomes [21].

2.2 CUMULATIVE ASSESSMENTS

There is consensus on the benefits of cumulative assessments for comprehensive learning outcomes. Cumulative assessments, employed across diverse educational settings, evaluate students' overall understanding and retention of knowledge and skills acquired over time. This approach offers advantages for assessing comprehensive learning outcomes, as it generally covers material from multiple courses, facilitates a comprehensive measure of learning progression, and encourages deeper understanding [22]. Cumulative assessments help ensure minimal competency, identify knowledge gaps, and foster accountability for cumulative knowledge and skills [22]. Various forms of cumulative assessments, such as exams or projects, prompt students to integrate knowledge, enhancing critical thinking and problem-solving skills. Educators also benefit by gaining insights into the effectiveness of teaching strategies and curriculum design [23]. The implementation of cumulative assessment systems, supported by information technology, further enhances academic performance and training quality by monitoring students' progress and providing timely feedback to facilitate their learning journey [24].

2.3 COGNITIVE FATIGUE

Cumulative assessments, while valuable for evaluating overall learning, are often lengthy, requiring students to spend significant time answering questions and completing tasks. This extended time-on-task can lead to cognitive fatigue, a state of mental weariness that negatively impacts performance across various domains [25]. As the test duration increases, students tend to experience a noticeable decline in test scores, further emphasizing the connection between cognitive fatigue and academic outcomes. This relationship is significant, as cognitive fatigue remains a powerful predictor of performance, even when other established factors are accounted for [26].

The subjective experience of fatigue intensifies with prolonged engagement in tasks, as shown by research indicating that cognitive fatigue increases with task duration, regardless of cognitive load [1]. This underscores the importance of considering both time constraints and the availability of cognitive resources when assessing the impact of cognitive fatigue [27]. Furthermore, the influence of cognitive fatigue extends beyond academic performance to affect decision-making processes, with research indicating a relationship between cognitive fatigue, response bias, and brain activation patterns [28].

Interestingly, some studies suggest that longer testing sessions or additional exam items may have positive effects, challenging the traditional view that cognitive fatigue always impairs performance. Incorporating more items into assessments has been linked to improved scores and enhanced performance [29]. Additionally, there is evidence suggesting that cognitive fatigue may even facilitate procedural motor sequence learning, pointing to a more complex relationship between fatigue and skill acquisition [30]. These findings suggest that cognitive fatigue's impact on learning and performance may depend on the context and type of task, necessitating further exploration of its nuanced effects on cumulative assessments.

2.4 CHUNKING OF EXAMS

While there is consensus on the benefits of cumulative assessments for comprehensive learning outcomes, there is divergence within the educational community regarding the methodology, particularly in the debate between an all-at-once approach versus a 'chunking' strategy, which segments the assessment over time. The chunking approach has been recognized as a beneficial strategy in both testing and learning contexts, as supported by academic research and institutional practices. Drexel University utilizes chunking as a testing accommodation, enabling students with variable conditions to complete exams in segments within a designated timeframe, ensuring exam integrity and fairness [31]. Similarly, the University of Massachusetts Amherst Center for Teaching and Learning highlights that breaking down complex information into smaller parts aligns with the brain's natural processing methods, reducing cognitive overload and improving knowledge retention [5].

Empirical studies further validate the effectiveness of chunking. In working memory tasks, chunking enhances information encoding and retrieval by reducing the load on working memory [32]. In the online learning domain, students preferred chunk-style videos over traditional lectures, with the former leading to higher engagement and improved learning outcomes [4]. Chunking course materials has been found to improve exam pass rates, suggesting that this strategy, when integrated with other teaching methods, could boost student success rates [33]. Additionally, chunked exams in large general education courses allow students to achieve learning outcomes in stages [34]. Chunking case studies into videos significantly enhanced student learning outcomes in microbiology [35]. Finally, breaking learning into manageable chunks in Biomedical Science programs supports student performance, leading to increased engagement and improved grades [36].

3 BACKGROUND

We explored the effect of chunking on cumulative exam performance in a hybrid online and hands-on rocketry course. Two groups of students participated in the course. Group A received a single final cumulative assessment (unchunked) while group B received a chunked final cumulative assessment. The course structure is shown in Figure 1 and Table 1. Figure 1 shows the course structure that is common for both groups A and B, while Table 1 illustrates the different cumulative assessment structure between groups. We implemented a structured approach to assess student engagement and learning outcomes in technical content. The number of questions associated with the surveys and technical quiz assessments is given in Table 2.

The course is structured as follows. It begins with an initial pre-content evaluation of student self-efficacy and interest through surveys. Then students engage with online technical content. Each week is a different technical topic and there are four total technical topics (Introduction to Rocket Hardware, Fundamentals of Rocket Design, Modeling Rocket Mechanics, and Flight Analysis). We administer a pre-content technical quiz at the beginning of each week to gauge baseline understanding before students engage with that week's content. At the end of the week, after engaging with the content, students complete a mid-content technical quiz that is identical to the pre-content quiz. We repeated the same self-efficacy and interest surveys after the online technical content at about the mid-content point. Students then complete the hands-on part of the course that is building, launching, and analyzing data from a model rocket. The final cumulative exam is administered at the end of the course and, up to this point, both groups A and B have experienced the same structure and timing and presentation of content.

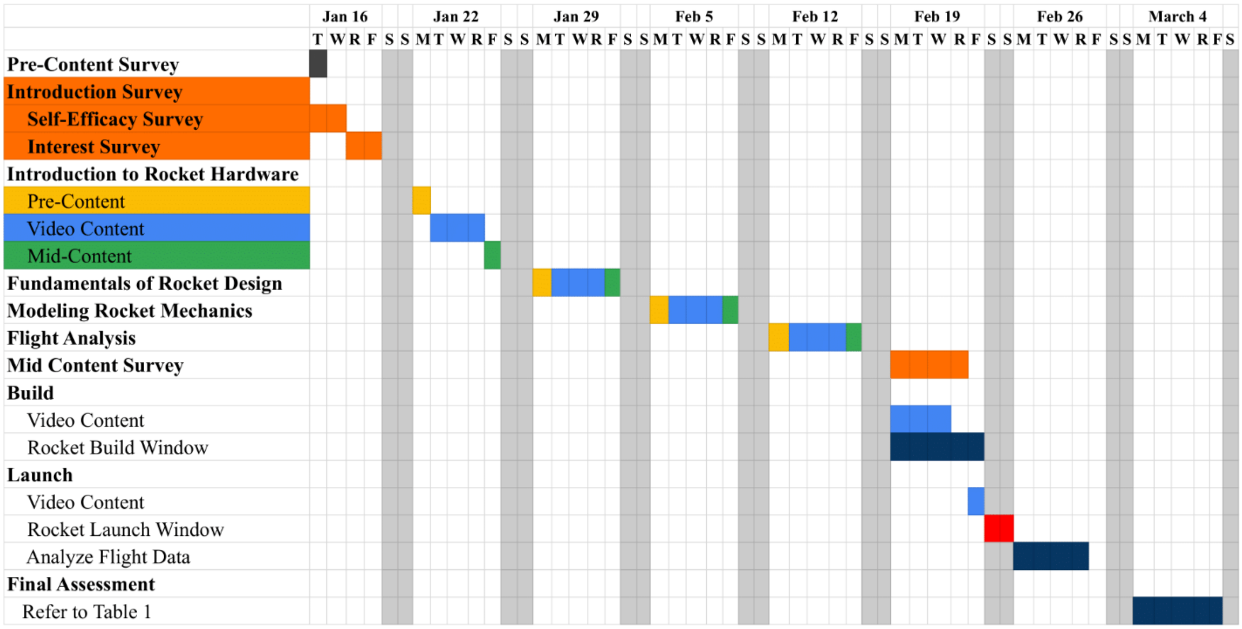


Figure 1: Common Course and Assessment Structure for Groups A and B

The final cumulative assessment consisted of all course evaluation materials. Specifically, students were asked to again complete the self-efficacy and interest surveys, along with the technical quizzes for all four technical content sections Table 2. However, the structure and timing of the different surveys and quizzes within this cumulative assessment was different for group A and B. Table 1 illustrates the different cumulative assessment structure between groups. Group A was tasked with a cumulative final exam with a flexible timeline and forced sequential completion of surveys and quizzes. Specifically, students were forced to complete the interest survey, self-efficacy survey, and technical quizzes 1-4 in that order. Note that the lengthy surveys are first. Further, the student was required to complete all these assessments in one sitting because progress could not be saved. The students of Group A had flexibility to complete the cumulative exam task at their discretion any time during the week. Group B was tasked with a "chunked" cumulative assessment. As Table 1 shows, different sections of the cumulative assessment were distributed across the week and required to be completed on specific days. Specifically, the cumulative assessment was divided into three parts: technical quiz 1-2, technical quiz 3-4, and self-efficacy and interest survey. Note that the lengthy surveys are at the end of the cumulative assessment for group B.

Table 1: Final Cumulative Assessment Structure

Group	Mon	Tues	Wed	Thurs	Fri
A	Interest, Self-Efficacy Survey, Tech Quizzes 1-4 (in order, at students' discretion)				
B		Tech Quiz 1-2	Tech Quiz 3-4	Interest and Self-Efficacy	

Table 2: Assessment Type and Number of Questions

Assessment	Number of Questions	Type
Self-Efficacy Survey	39	7-point Likert scale
Interest Survey	25	5-point Likert scale
Introduction and Rocket Hardware	7	Multiple Choice
Fundamentals of Rocketry	6	Multiple Choice
Modeling Rocket Mechanics	6	Multiple Choice
Analysis	5	Multiple Choice

4 LOGIC MODEL AND THEORY OF CHANGE

Our Logic Model, shown in Table 3, delineates what we believe based on the literature review citations, to be the causal connections between the exam structure and the intended outcomes (student performance and testing fatigue). It categorizes two cohorts: Group A, tasked with an 'unchunked' cumulative exam, and Group B, tasked with a chunked exam. Our logic model posits that an unchunked exam correlates with diminished student scores and increased testing fatigue, attributed to inadequate preparation and flexible completion schedule. Conversely, chunked exams yield elevated scores and enhanced readiness, compelling students to complete exams within designated timeframes. The logic model posits that the implementation of chunked exams leads to enhanced student performance and diminished testing fatigue compared to unchunked exams. A restructured exam could lead to better learning outcomes and less fatigue, providing a clearer and more accurate assessment of how students engage with and understand the course content.

Table 3: Logic Model

Group	Inputs	Outputs	Outcomes/Impact
A	• "Unchunked" Final Exam • Students finish at their own pace • All final exam sections due on the same day	• Lower student scores • Less accurate representation of student knowledge	• Superficial understanding of content • Shorter knowledge retention
B	• "Chunked" Final Exam • Forced pace to finish on specific days • Final exam sections due on different days	• Students are less stressed and review materials • Higher student scores • More accurate representation of student knowledge	• Deeper understanding of content • Longer term knowledge retention

5 METHODS

5.1 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 effect of assessment structure on perceived outcomes of participants in this course.

5.2 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. The course was graded based on participation. Students received two hours of credit for completing all surveys and quizzes (all participants in this study received full credit). This level of course credit is insufficient to fulfill technical elective requirements, which typically require three or more credits. 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 complemented by qualitative surveys, was designed to glean nuanced insights into student experiences, potentially uncovering hidden gaps or strengths that might elude detection through surveys alone.

5.3 DATA ANALYSIS

Survey responses and technical quiz scores were acquired and analyzed. 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, with a significance level (alpha) set at 0.01. We first conducted a Shapiro-Wilk test to assess the normality of our data. For normally distributed data, we applied a paired T-test to compare pre- and post-assessment scores within each group, which allowed us to evaluate the progress of individual students. For non-normally distributed data, we utilized the Wilcoxon signed-rank test as a non-parametric alternative to the paired T-test. To compare the two groups' performance, we used an independent T-test (for normally distributed data) or the Mann-Whitney U test (for non-normally distributed data) to determine if the differences are significant.

This difference in distribution is prevalent in types of assessments. These can be broken up into two parts, Interest and Self-Efficacy Surveys, and technical quizzes. Both the surveys are assessed on a Likert-Scale, with the difference being that the Interest is on a 5-point scale and the Self-Efficacy is on a 7-point scale. The technical quizzes follow a traditional grading format ranging from 0-100%.

6 RESULTS

6.1 STUDENT DEMOGRAPHICS

Details and demographics of Groups A and B are shown in Table 4. Group A comprised 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 college. 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 comprised twenty-five undergraduate students. This group consisted of 40% male and 60% female students. In terms of ethnicity, 12% identified as Hispanic or Latino/a, while 88% were not. Regarding race, 36% identified as Asian, and 64% as white. Furthermore, 60% were first-year students, and the fields of study included mechanical engineering (28.0%) and astrophysics (12.0%), with physics (4.0%) also represented. Similar to Group A, early-stage college participants were favored in selection, resulting in mostly freshmen and sophomores (84%), with a smaller proportion being juniors and seniors (16%).

Table 4: Student Demographics

	Group A		Group B	
Categories	**n**	**Percent %**	**n**	**Percent %**
Total	32	100	25	100
Gender				
Female	11	34.4	10	40.0
Male	20	62.5	15	60.0
Prefer not to say	1	3.1	0	0.0
Ethnicity				
Do not wish to provide	1	3.1	0	0.0
Hispanic or Latino/a	6	18.8	3	12.0
Not Hispanic or Latino/a	25	78.1	22	88.0
Race (Multiple selections allowed)				
American Indian or Alaska Native	1	3.12	0	0
Asian	19	59.9	9	36.0
Black or African American	2	62.5	0	0
White	12	37.5	14	64.0
Do not wish to provide	1	3.1	0	0
Year in College				
1	19	59.4	15	60.0
2	9	28.1	6	24.0
3	3	9.4	3	12.0
4	1	3.1	1	4.0
Degree Program				
Ag. & Biomed. Eng.	1	3.1		
Astrophysics			4	16.0
Astronomy			3	12.0
Civil Eng.	1	3.1		
Chem. Eng.			2	8.0
Comp. Eng.			1	4.0
Comp. Sci.	1	3.1		
Elec. & Comp. Eng.	2	6.2	2	8.0
Eng. Mechanics			1	4.0
Eng. Undeclared	1	3.1		
Ind. & Ent. Sys. Eng.	4	12.5	1	4.0
Mat. Sci. & Eng.	2	6.2		
Mech. Eng.	9	28.1	7	28.8
Math	2	6.2	1	4.0
Nuclear Eng.			1	4.0
Physics	8	25.0	2	8.0
Business	1	3.1		

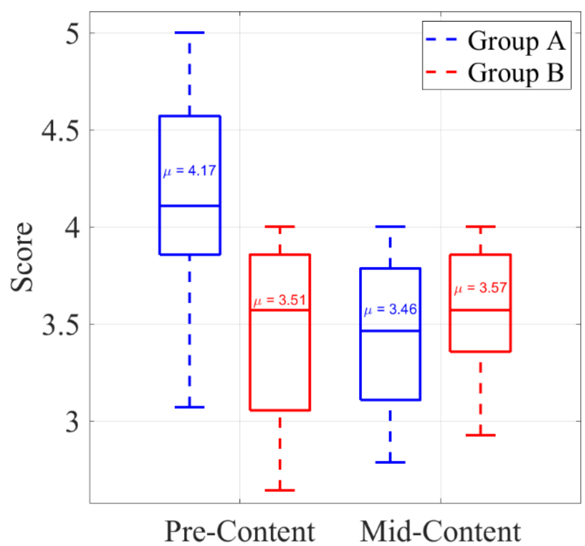
6.2 ASSESSMENTS PRIOR TO CUMMULATIVE ASSESSMENT

The assessments prior to the cumulative final assessment consisted of the pre-content and mid-content surveys and technical quizzes. Overall, analysis of these assessment results suggests that there are no significant differences in the initial technical knowledge level or growth in knowledge between groups A and B. Further, each group started the course with similarly high levels of interest and self-efficacy, which remained high in the mid-content assessment. Group B consistently reported significantly higher levels of self-efficacy throughout the duration of the course.

6.2.1 INTEREST AND SELF-EFFICACY

Figure 2 and Figure 3 together provide a detailed analysis of the changes in interest and self-efficacy for both groups. The hypothesis tests were conducted to compare the change in self-efficacy and interest for both groups during the pre-content and mid-content phase. The null hypothesis for these tests posits that there is no significant change in self-efficacy and interest levels within the groups between pre- and mid-content assessments. The associated p-values from these tests provide the basis for accepting or rejecting this null hypothesis, thereby indicating whether the observed changes in self-efficacy and interest are statistically significant for each group.

In Figure 2, the left plot illustrates a significant decline in interest for Group A from the pre- to mid-content assessment. This result is statistically significant as shown by the data in Figure 3, which quantifies the change in the average score between pre- and mid-content. In contrast, Group B interest levels remain relatively unchanged. The right plot in Figure 2 shows a significant increase in self-efficacy from the pre- to mid-content, which is further supported by the statistical findings in Figure 3. This evidence indicates growth in self-efficacy for both Group A and Group B, suggesting that students' confidence in their abilities to succeed in the course is enhanced as they progress through the material. The combined data from Figure 2 and Figure 3 reveal a nuanced dynamic between interest and self-efficacy within the course content between the two groups. While Group A reports a notable decline in interest, the self-efficacy of both groups significantly increased, suggesting that course content may enhance students' confidence in their academic abilities, even as their interest levels decline.



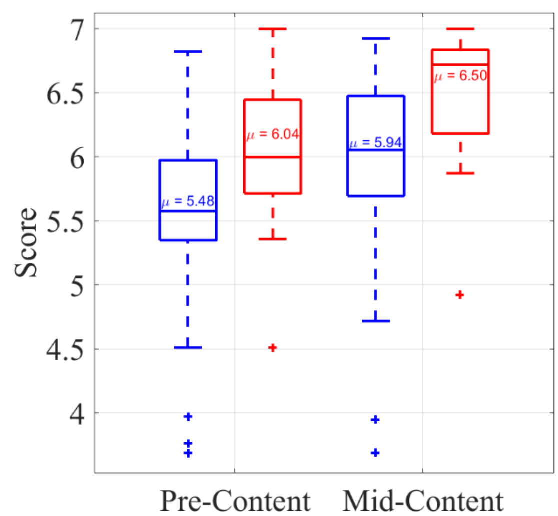


Figure 2: Interest (left, 5-pt scale) and Self-Efficacy (right, 7-pt scale) prior to the final assessment.

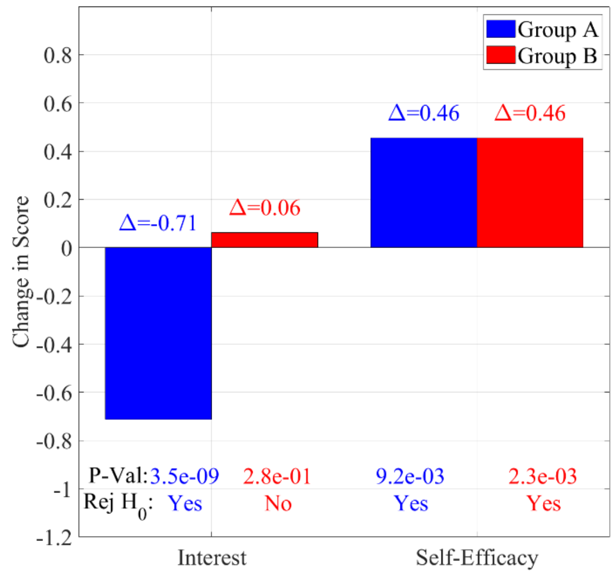


Figure 3: Changes in Interest and Self-Efficacy from Pre-Content to Mid-Content

6.2.2 TECHNICAL QUIZZES

Figure 4 presents a comparison of the distribution of technical quiz scores between the groups for each technical content module, with the mean scores indicated inside each box. During the pre-content phase (left plot in Figure 4), there is a general trend of increasing mean scores across all modules for both groups, except for a decrease in module 3. It is interesting to note that module 3 is more computationally intense, and this may explain the lower mean scores observed. For Group A in module 2, all scores are above the lower quartile, suggesting a higher baseline performance. Group B's scores for modules 2 and 3 are more equally distributed about the median, as indicated by the even spacing between the lower and upper quartiles. Group A's scores for modules 1 and 4 also display this balance, while module 3 shows a wider spread of scores below the median, reflecting greater variability in lower performance.

=

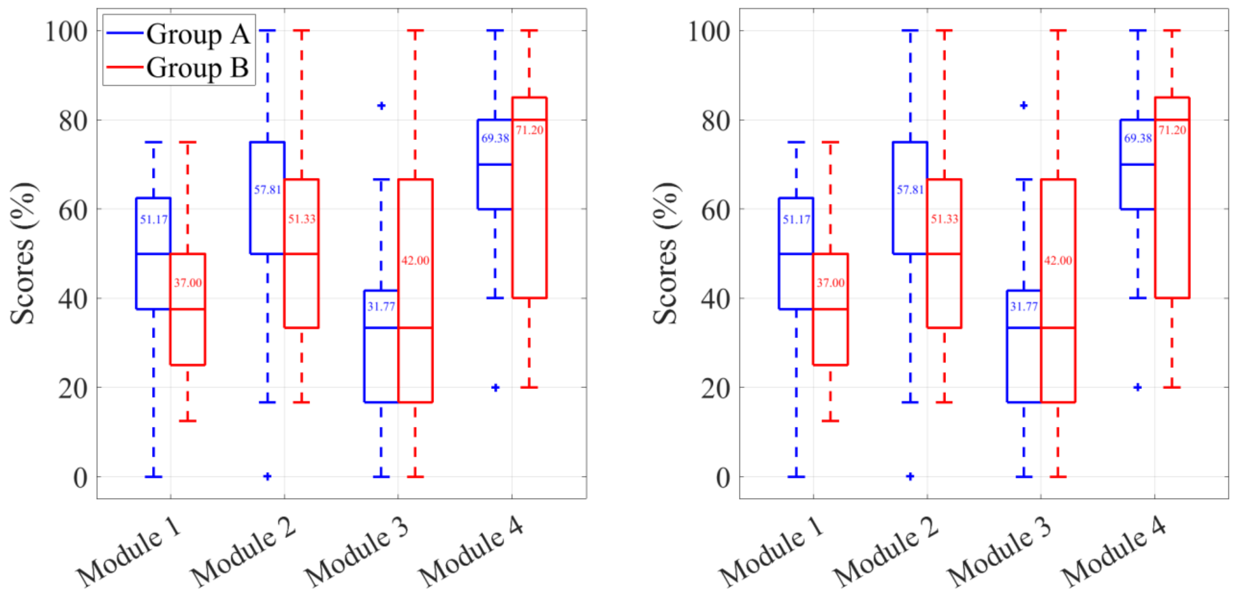


Figure 4: Group A and B Technical Quiz Scores for each Technical Module pre-content (left) and mid-content (right)

In the mid-content phase, as depicted in the right plot of Figure 4, we observe an overall increase in scores and performance for both groups compared to the pre-content assessment phase (left plot). For Group A, module 1 scores are evenly distributed between the lower and upper quartiles. However, for modules 2 through 4, the scores are predominantly clustered towards the upper end of the spectrum. This is indicated by the absence of scores above the upper quartile and, as queried, the median (represented by the middle bar on box and whisker plots) aligns with the upper quartile edge. This phenomenon occurs when the data has certain characteristics, such as outliers or when it is skewed, causing the median to coincide with the quartile boundary. It could imply either a robust grasp of the content or a potential ceiling effect in the assessment's ability to distinguish among top-performing students.

Group B's performance varies, with modules 1 and 3 displaying a lower median than modules 2 and 4. Notably, for module 4, Group B's scores are remarkably uniform, with all scores nestled within the lower and upper quartiles, indicating a consistent performance across the group. Group A presents negative outliers in modules 1 and 2, highlighting areas where a subset of students scored substantially lower than their peers. Conversely, Group B exhibits a broader spectrum of performance in module 4, with one positive outlier and two negative outliers, suggesting a more diverse range of outcomes within that module.

Figure 5 shows the change in mean scores across the modules between the groups, including the p-value and hypothesis decision. For both groups and all modules, the average quiz score increased from pre-content to mid-content assessment. The average quiz performance increased most for module 1, which is the first module and at the beginning of the course. Group B's performance improvement decreased almost linearly from module 1 to 4 throughout the course. For all modules, group B average quiz score increased more than group A. Group B performance increased on average 19.8% for each module, while group A performance increased only 8.5%.

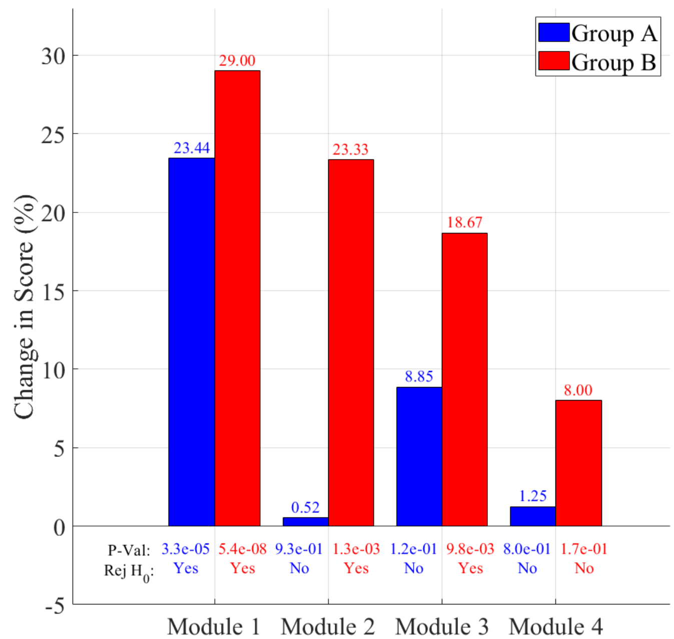


Figure 5: Change in Average Technical Quiz Score from Pre- to Mid-Content for each of the Four Modules

Scores between group A and B were compared using a two-sided hypothesis test. Table 5 shows these comparisons. The null hypothesis asserts that there are no differences in the mean scores of the two groups across modules and surveys, suggesting that any observed differences are merely due to random variation. The alternative hypothesis contends that there are significant differences between the groups.

Table 5 presents the p-values from our statistical tests alongside decisions on the null hypothesis. For the interest survey in the pre-content phase, the p-value is 4.4e-07, which leads to the rejection of the null hypothesis, confirming that Group A had a significantly higher level of initial interest. In contrast, the mid-content phase showed no significant difference in interest levels between the groups, with a p-value of 2.6e-01. Group B reported significantly higher levels of self-efficacy at both the pre-content and mid-content phases, with p-values of 1.1e-03 and 3.2e-03 respectively. Technical quiz scores showed no significant difference at either phase, supporting the assertion that the groups are similar overall. For all the technical module scores, the p-values do not lead to the rejection of the null hypothesis for both the pre-content and mid-content comparisons. This suggests that, in terms of technical quiz performance, there are not significant differences between group A and B.

Table 5: Hypothesis Results Between Groups

	Pre-Content		**Mid-Content**	
Module/Survey	**P-value**	**Reject Null**	**P-value**	**Reject Null**
Interest	4.4e-07	Yes	2.6e-01	No
Self-Efficacy	1.1e-03	Yes	3.2e-03	Yes
Module 1	4.6e-01	No	7.3e-01	No
Module 2	8.9e-01	No	6.7e-01	No
Module 3	8.4e-01	No	9.1e-01	No
Module 4	9.4e-01	No	1.4e-02	No

6.3 FINAL ASSESSMENT RESULTS

Figure 6 left presents results for each module within the technical knowledge portion of the final assessment. For module one, the mean scores for Group A and Group B are similar, at approximately 72% and 71%, respectively. However, Group B outperformed Group A on the remaining three modules by on average 13.1%. Figure 6 right illustrates the cumulative final assessment technical module score distributions for Group A and Group B, revealing distinct differences in performance. Group A achieved an average score of 61.6, with a wide range of scores stretching from approximately 35 to 90 percent. The distribution for Group A is skewed towards higher scores, as indicated by an upper quartile that is substantially larger than the lower quartile, suggesting that a significant number of students scored in the higher score range. In contrast, Group B's average score was notably higher at 73.9, and their score distribution was more balanced, with nearly even upper and lower quartiles, which points to a more uniform distribution of scores across the group.

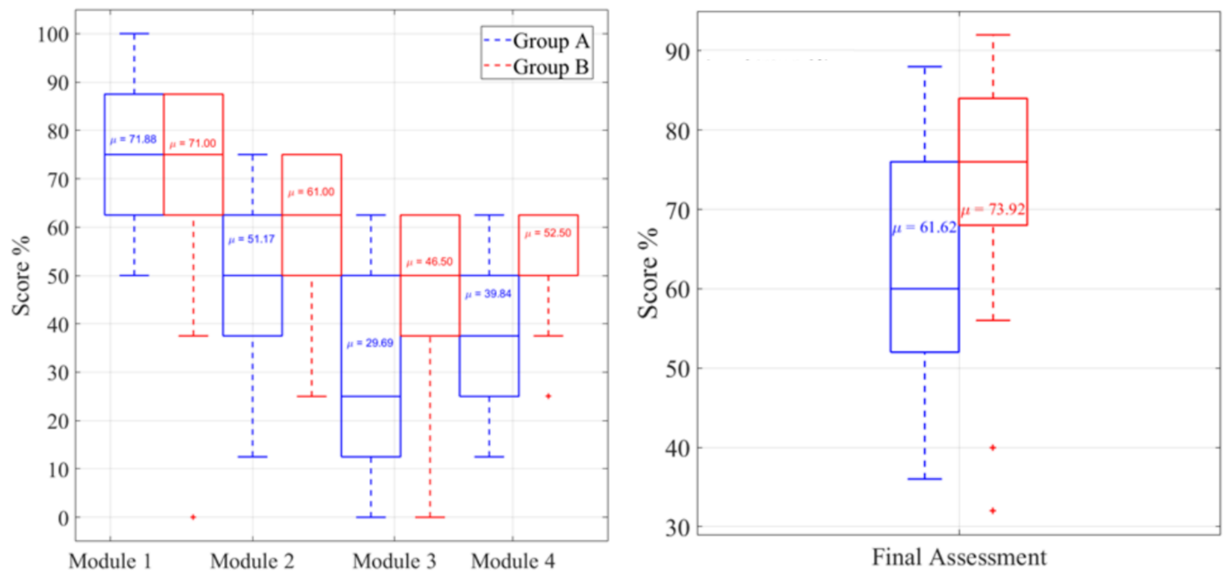


Figure 6: Final Assessment Technical Quiz Scores by Module (left) and Cumulative (right)

Figure 7 shows the interest and self-efficacy survey results. The same hypothesis testing was employed for this cumulative assessment to determine significant differences between groups and those results are reported in Table 6. Group A reported higher interest levels, while Group B reported higher self-efficacy, with significant differences between the groups. On interest surveys, Group A's scores ranged from 3 to 5, with a mean of 4.17 and a higher upper quartile, indicating more students reported high interest. Group B's interest scores ranged from 3 to 4, with a lower mean of 3.59 and a higher lower quartile, reflecting lower interest levels. Notably, Group B's interest levels remained unchanged throughout the course, maintaining an average of 3.59. (Figure 7) as at the mid- and pre-content assessments, 3.57 and 3.51 (Figure 2), respectively. On the final assessment, group A reported a return to their pre-content interest level. Group A’s average final interest level is 4.17 (Figure 7) and is the same as the pre-content level, 4.17 (Figure 2). This is an increase from the average mid-content interest level of 3.46 (Figure 2).

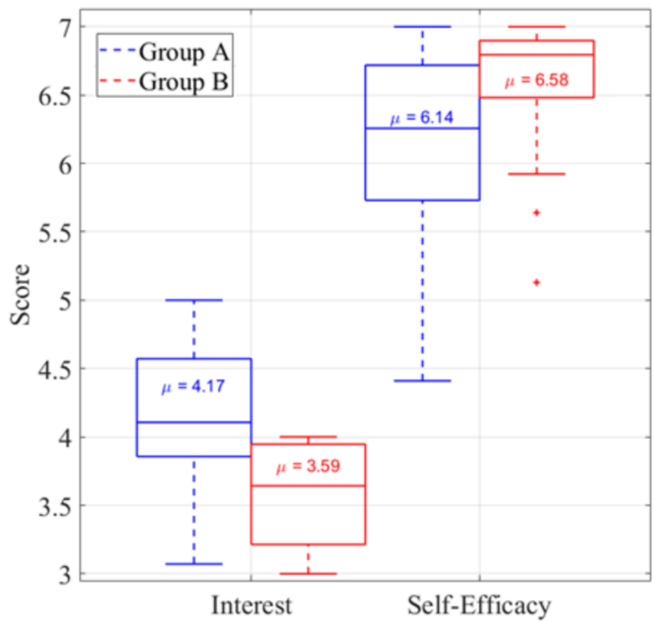


Figure 7: Final Assessment Results from Interest (5-pt scale) and Self-Efficacy (7-pt scale) Surveys

Table 6: Final Assessment Score Comparison Hypothesis Results

Module/Survey	**P-value**	**Reject Null**
Interest	1.1e-13	Yes
Self-Efficacy	2.5e-06	Yes
Technical Quiz Cumulative	6.43e-03	Yes

6.4 STUDENT ENGAGEMENT THROUGHOUT THE COURSE AND FINAL ASSESSMENT

6.4.1 ONLINE VIDEO CONTENT ACCESS

We investigate student engagement during the course and during the final assessment week by tracking and analyzing course online video access and viewership. Specifically, we report here the number of times a video was viewed and the average duration of each viewing. Figure 8 reports these engagement metrics for the four-week period during which the technical modules online video content was assigned (weeks of Jan 22 to Feb 12 in Figure 1). Figure 9 reports the same engagement metrics for the week of the cumulative final assessment (week of Mar 4 in Figure 1).

Figure 8 is a plot of the total views for each video for each group before the final assessment week. There are a total of 31 videos available sequentially across the four technical modules (M1, M2, M3, M4). A common trend is observed where views decrease as the video number increases. In other words, views are initially high for the first video within a module but tend to decrease as the module progresses. Further, views tend to decrease across all videos regardless of the module. This pattern suggests that students are most engaged at the beginning of the course and at the beginning of a new topic (module), with interest waning as they become more familiar with the content. However, Module 3 (M3) deviates from this trend, with views increasing towards the end of the module. This increase in engagement is likely attributed to the last video in M3, which covers a skill directly related to a computationally difficult technical quiz question (related to modeling and plotting a rocket's trajectory).

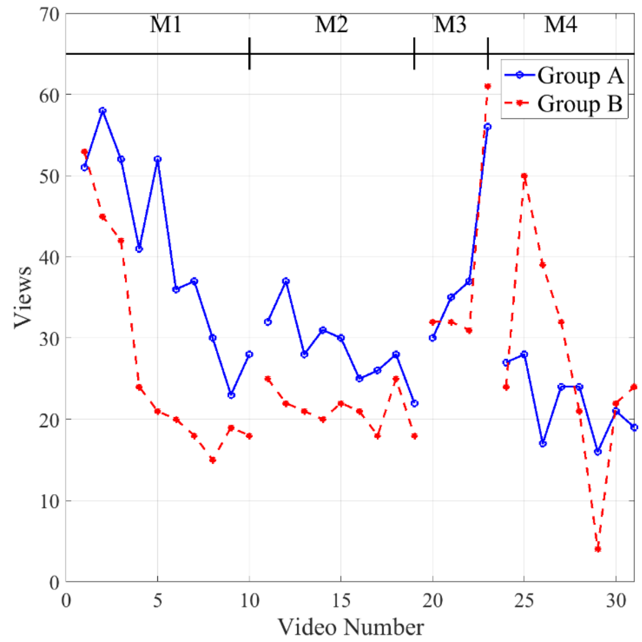


Figure 8: Number of Views before final assessment

Due to the high number of views for the final video in Module 3 (video 23), we conducted a deeper analysis of its statistics. We found that it had the longest view duration of all the videos, averaging 350 seconds, compared to an average view duration of 200 seconds for the other videos. However, despite having the longest view duration, only 30% of the entire 12-minute video was viewed on average, the smallest fraction of any video. In contrast, the other videos had an average fractional viewing duration of 65%. This suggests that students engaged with the video 23 just long enough to gather the necessary information to complete two specific quiz questions in Module 3, resulting in lower overall view percentage despite its length.

Introductory videos in Modules 2, 3, and 4 consistently showed the longest view durations within their respective modules with an average of 35 and 34 seconds for Group A and B respectively. While the rest of the videos within a module averaged 30 and 25 seconds for Group A and B respectively. This suggests that while students engage more with the introductory videos, they may spend less time on the actual technical content. The first video of Module 1, however, stands out with a notably high view duration, highlighting the significance of early engagement. Figure 9 (left) shows video viewing during final assessment week. Group A shows minimal interaction with the course videos, with negligible views across most modules. Interestingly, there are zero views for the video 23 in Module 3, despite its direct relevance to a computationally intensive question on the final. Group B's engagement is higher and more consistent, with multiple views in each module. A notable peak occurs with video 5 of Module 1, which accumulated a total of 15 views. For Modules 2 through 4, Group B's views are generally within the range of 1-6 views, indicating a steady but moderate engagement with the content during finals week. We note that group B also did not engage with video 23.

Figure 9 (right) presents a breakdown of the average view duration per video. The same trend observed in Figure 9 left is apparent in Figure 9 right with Group A showing almost no engagement, resulting in negligible view durations. However, for the few videos that Group A did engage with, they played the entire video, as indicated by the view durations matching the video lengths for video 5 and 18. This could imply that Group A students were not actively engaging with the content, but rather letting the videos play through to the end while they were engaged in another task. On the other hand, Group B's view durations range from 0 to 200 seconds. None of the view durations for Group B during finals week correspond to the full length of the videos, which may suggest that students are selectively watching specific segments of the videos to review particular topics or find answers for the final exam.

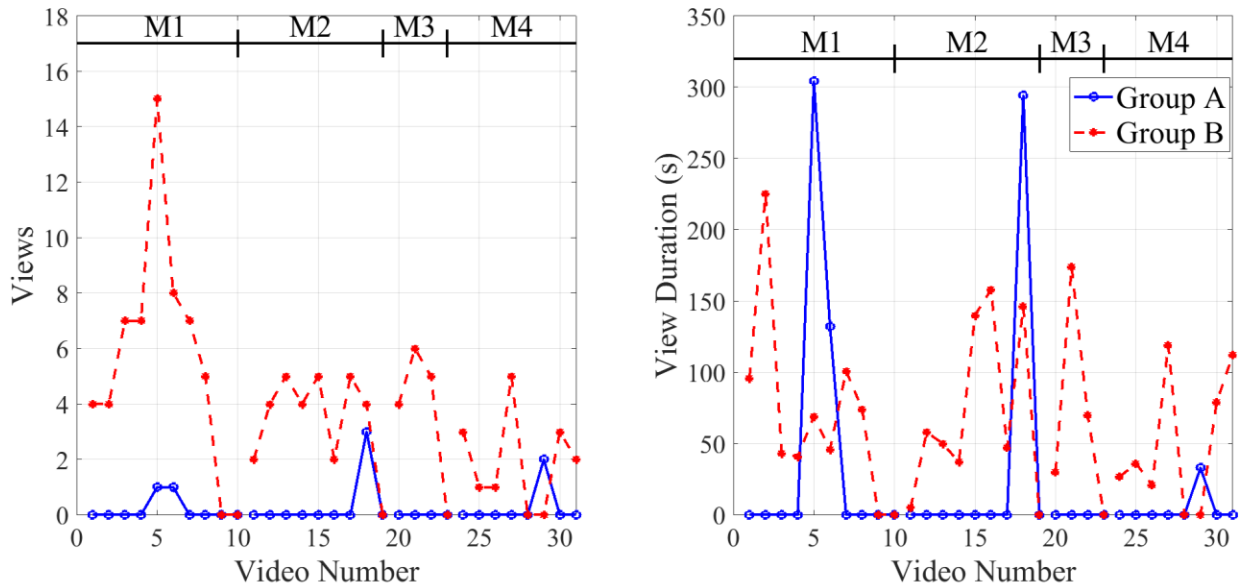


Figure 9: Views (left)and View Duration (right) During Final Assessment Week

Table 7 presents the results of statistical hypothesis tests conducted to compare the view counts and view durations between the two groups throughout the course and specifically during finals week. The null hypothesis posits that there is no significant difference in the view counts and view durations between the two groups, both during the course and the final assessment week. The alternative hypothesis is that a difference does exist. Before the final assessment week, the resulting p-values for the views (8.4e-1) and view duration (9.1e-1) exceed the predetermined alpha level of 0.01, indicating insufficient evidence to reject the null hypothesis. Consequently, we cannot conclude that there is a statistically significant difference in the engagement metrics between the groups before the final assessment week. However, during the final assessment week, the analysis yields p-values of 9.1e-09 for views and 1.5e-06 for view duration, both significantly lower than the alpha threshold. This provides strong evidence to reject the null hypothesis in favor of the alternative, suggesting that Group B's engagement with the course content was indeed higher than that of Group A during the final assessment week.

	P-Value	Reject Null
Views before final	8.4e-1	No
View Duration before final	9.1e-1	No
Views during final	9.1e-09	Yes
View Duration during final	1.5e-06	Yes

Table 7: View and View Duration Hypothesis Results

6.4.2 FINAL ASSESSMENT AND SUBMISSION METRICS

We further investigate student engagement during the cumulative final assessment week by tracking and analyzing submission metrics for the final assessment. Specifically, we report the number of submissions each day, the average technical quiz score associated with submissions on a particular day, and the percentage of the student-exam completed. To clarify, for the 32 Group A students, everyone must complete the entire exam in one sitting. Therefore, one student finishing the exam is equivalent to 1/32 = 3.1% of the total student-exam. In contrast, Group B students, with exam chunking, are required to complete a section of the exam each day. For instance, if five of the 25 students complete the first of three exam sections, then (5/25)x(1/3) = 6.6% of the entire student-exam has been completed.

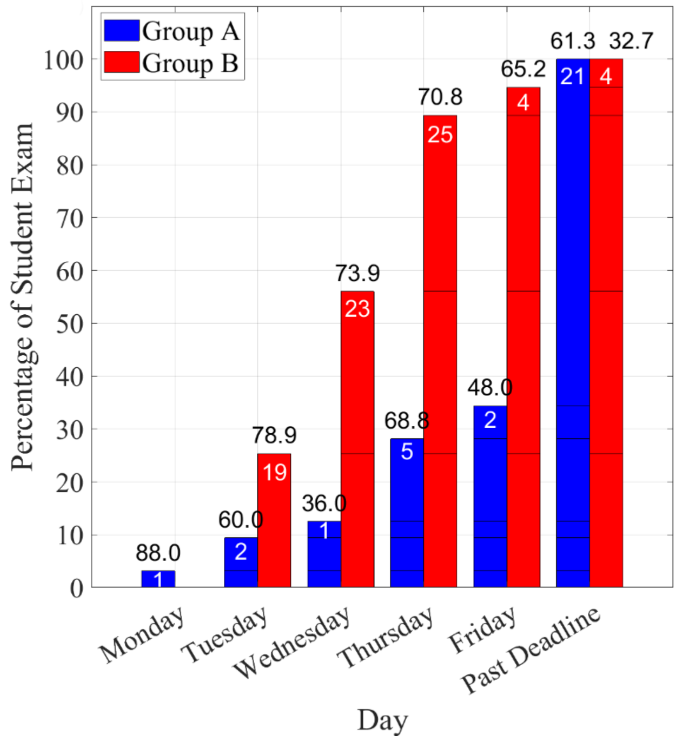


Figure 10: Final Assessment Week Submission Metrics

Figure 10 shows the submission metrics for the final assessment week. The white number inside the bar is the number of submissions on that day. The black number above the bar is the average score for the submissions that day. The height of the bar is the cumulative percentage of student exams submitted. For example, on Thursday of the final assessment week, there were five group A submissions with an average technical quiz score of 68.8% and these submissions brought the overall student-exam percentage up to 28% (1+2+1+5=9 out of 32). Also on Thursday, there were 25 group B submissions with an average technical quiz score of 70.8% and these submissions brought the overall student-exam percentage up to 89%. It is important to note that, for group B, Tuesday and Wednesday were dedicated to technical quizzes, whereas Thursday's session was supposed to be Interest and Self-Efficacy surveys only. So, there should have been no technical quiz scores on Thursday. However, there are technical quiz scores reported for group B on

Thursday, and later, because students completed and submitted the technical quiz after the deadline.

For Group A, who took the unchunked exam, submissions fluctuated throughout the week. One student completed the exam on Monday, achieving a mean score of 88. By Tuesday, the number of submissions increased to two, but the mean score dropped to 60. A surge in submissions occurred on Thursday, with five students completing the exam, leading to an improved mean score of 68.8. Two additional submissions were made on Friday with a mean score of 48. Notably, 21 students submitted their exams after the Friday deadline, with an average score of 61.3. Group A's completion level reached 100% after the deadline, indicating a spread-out submission pattern over the week. There does not appear to be any clear trend between when the final assessment is submitted and the resulting technical quiz score. But the data do show that 60% of the group A students with an unchunked exam submitted the final assessment after the deadline.

In stark contrast, Group B, with a chunked exam format wherein exams and surveys were distributed across Tuesday to Thursday, had a more uniform submission pattern. The percentage of student-exam submissions for Group B were 25.3%, 56.0%, and 89.3%for Tuesday, Wednesday, and Thursday respectively, which is slightly below the ideal 33%, 66%, and 100%, respectively. By Friday, the percentage of student exams completed had marginally risen to 94.7%, with the remaining submissions arriving further after the deadline. In contrast to group A, the percentage of student-exams submitted after the deadline was about 10% for group B. The structured submissions of group B contrasts with Group A's delayed and later completion timeline and suggests that a chunked exam format may encourage more punctual and timely submissions.

We observe that Group B's mean technical quiz scores began high on Tuesday at nearly 79%, which included only part 1 of the chunked exam (modules 1 and 2). The mean score decreased to about 74% for the Wednesday submissions, which included both the second part of the technical quiz and late submissions from part 1. The late technical quizzes received on Thursday, Friday or beyond had a mean of about 70%, 65%, and 33%, respectively. These data suggest that later submissions have lower scores for group B, which is to be contrasted with group A where no trend was discernible.

7 DISCUSSION

The primary aim of our data analysis is to meticulously identify and understand any significant disparities in the behavior of the two groups both before and during the final examination period. By closely examining these engagement patterns, we aim to establish a baseline of comparability between the groups. This is a crucial step in ensuring that any observed differences in performance can be confidently attributed to the impact of the modified final assessment structure, rather than pre-existing variations in study behavior or content interaction.

7.1 INVERSE RELATIONSHIP BETWEEN SELF-EFFICACY AND INTEREST

One aspect of our study explored the relationship between self-efficacy, interest, and engagement patterns in two groups of students throughout a course, focusing on changes before and during the final examination period. A key finding was that self-efficacy increased for both groups from pre- to post-assessments, with Group B consistently reporting higher self-efficacy than Group A. Meanwhile, interest levels remained relatively stable, with Group A showing higher initial interest that fluctuated mid-course but returned to its original level by the end.

The rise in self-efficacy observed across both groups aligns with previous research demonstrating that self-efficacy tends to increase as individuals gain experience and develop confidence in their abilities [37]. However, while confidence improved, interest levels showed minimal change, suggesting that self-efficacy alone may not necessarily drive heightened interest, which is supported by research indicating that while self-efficacy can predict academic achievement, it does not always correspond to shifts in interest [38]. This highlights a more nuanced relationship between these variables.

Interestingly, our data revealed an inverse relationship between self-efficacy and interest. Group B, despite consistently higher self-efficacy, reported lower interest compared to Group A, which maintained a higher interest level despite lower self-efficacy. This suggests that students may engage more in areas where they feel less competent, perceiving greater potential for growth [39]. This could explain why Group B's high self-efficacy did not lead to a significant increase in interest—confidence might reduce the need for exploratory learning.

Engagement patterns between the groups further illustrated this dynamic. Group A's engagement declined as the course progressed, particularly during finals week, potentially due to their lower self-efficacy and fluctuating interest [37]. In contrast, Group B's sustained self-efficacy translated into consistent engagement, particularly during high-pressure periods like the final exam [38]. Higher self-efficacy is strongly linked to continued engagement and knowledge acquisition under challenging conditions.

Mastery experiences play a critical role in maintaining engagement by building self-efficacy [40]. This likely explains Group B's steady interaction with course materials, particularly the course videos, as their self-confidence allowed for more structured approaches to learning. While self-efficacy positively correlates with academic performance, it does not necessarily increase interest, indicating that fostering both requires distinct strategies [41].

Moreover, engagement metrics from our study suggest that Group A's initial enthusiasm waned due to insufficient self-efficacy to sustain it over time. A moderate level of self-efficacy is crucial to maintaining interest, as it drives curiosity through mastery experiences [42]. While Group A's engagement may have been fueled by their perceived competence early on, as reflected in their self-efficacy mean of 6.1, it ultimately wasn't enough to maintain consistent engagement throughout the course.

Conversely, Group B's consistently higher self-efficacy (mean of 6.6) likely contributed to their steady engagement, even during finals week, when they demonstrated higher interaction with the course videos. Hypothesis tests showed a statistically significant difference in engagement between the two groups during the final examination period, with Group B showing greater engagement. These findings suggest that while self-efficacy supports sustained engagement, educators should also focus on maintaining interest through varied and stimulating learning opportunities.

7.2 TECHINCAL QUIZ SCORES

The data indicates that students in both Group A and Group B began the course with comparable foundational knowledge, as evidenced by their closely matched pre-course technical quiz scores—Group A achieving a mean score of 52.5326% and Group B 50.3833%. This trend continued throughout the course, with mid-course quiz scores also reflecting a similar level of understanding; Group A scored a mean of 61.0% while Group B scored 70.1%. This suggests that both groups engaged with and comprehended the course material at a similar rate.

The lack of significant differences in quiz performance from the pre-course to the mid-course assessments implies that the educational content and teaching methods were equally effective and accessible for both groups. Both groups progressed similarly, as demonstrated by comparable video viewing metrics throughout the course.

However, a divergence in performance emerged during finals week, particularly evident in the final exam completion rates and scores. Group B demonstrated higher engagement and completion rates, as illustrated in Figures 9 and 10, along with their adherence to the chunked exam format. This indicates that differences in performance on the final exam may stem from factors such as study habits, time management, or exam preparation strategies rather than from initial knowledge gaps or variations in learning throughout the course. Therefore, the final assessment's all-at-once format may not effectively gauge students' knowledge, as both groups exhibited similar capabilities leading up to the final evaluation.

7.3 FINAL EXAM SCORES

The findings from the final exam hypothesis test, which yielded p-values of 9.1e-09 for finals week views amount and 1.5e-06 for finals week view duration, indicate that Group B's performance surpassed that of Group A. This is consistent with the observed engagement patterns where Group B demonstrated higher interaction with the course content during finals week, as shown in Figures 9 and 10. Furthermore, the hypothesis test for the technical quiz scores, conducted from the onset of the course through to the point after students had interacted with the online content and undergone assessment, showed no significant difference in performance between the two groups. This suggests that students from both groups, who share similar demographic characteristics, maintained equivalent levels of achievement throughout the course.

The key element that seems to have contributed to the higher final exam scores of Group B is the structure of the assessment itself. Given that the technical quiz scores were similar for both groups, the divergence in final exam performance points to the assessment design as a potential factor influencing the outcome. This leads to the inference that the design and format of the final exam were likely influential in creating the performance gap observed between the two groups.

This conclusion prompts a closer examination of the assessment methods and their potential impact on student performance. The fact that Group B's engagement with the course content was significantly higher during finals week suggests that the format of the final exam may have been more conducive to their study habits or preparation approach. Understanding how the structure of an assessment can affect outcomes is crucial for ensuring fair and accurate measures of student learning and can guide future improvements in test design to better reflect the abilities of all students.

7.4 ENHANCED STUDENT ENGAGEMENT WITH CHUNKED ASSESSMENT

The structure of assessments significantly influences student engagement and outcomes. Prior to the final assessment, both Group A and Group B exhibited similar levels of engagement, as indicated by their views and view durations. However, a noticeable shift occurred during the assessment week, as students in Group B, who participated in the chunked assessment, demonstrated increased engagement levels. This heightened engagement was evident through their enhanced video viewing during the assessment period and their timely submission of the final assessment.

The idea that Group B performed better due to alignment between their learning strategies and a chunked assessment structure is well-supported by research on chunking and its relationship to learning styles. Breaking study material into smaller, manageable chunks significantly improves retention and long-term memory, as students benefit from repeated exposure to information over time [43]. Additionally, chunked formats enhance focus and reduce attention lapses, suggesting that assessments divided into smaller segments might better suit learners who thrive on consistent engagement [44]. When the material is chunked, learners can break the task into achievable pieces, promoting sustained attention and deeper cognitive processing. This structured approach mirrors the "Study A Little A Lot" method, where regularly reviewing smaller segments of material helps reduce cramming and increase retention [45]. Collectively, these findings indicate that learners employing chunking strategies benefit from assessments that reflect these segmented approaches, which may explain Group B's success.

Pre-segmented material reduces cognitive load and improves learning efficiency, particularly for those with less prior knowledge [46]. When cognitive strain is minimized, students can process relationships between concepts more easily, leading to better performance. This idea is echoed by findings that students prefer chunking strategies over traditional didactic teaching methods due to the increased focus and improved long-term retention chunking provides [4] Chunking allows learners to form meaningful connections between pieces of information, enhancing their ability to recall and apply knowledge. Therefore, if Group B's learning style aligned with chunking, their performance was likely boosted by assessments that matched their study strategies, providing the cognitive support needed for higher engagement and understanding.

The fact that Group B's engagement with the course content was significantly higher during finals week suggests that the format of the final exam may have been more conducive to their study habits or preparation approach. Certain assessment formats align better with students' study strategies, particularly those that incorporate chunking [47]. Students exhibited a clear preference for assessments requiring deeper engagement, such as individual case studies, which facilitate chunking by breaking down complex information into smaller, manageable parts, thereby enhancing retention and understanding. Moreover, assessment formats involving critical analysis, such as essay writing and group discussions, resonated more with students who engaged in sequential processing [48]. While little correlation has been found between learning styles and academic performance, results highlight the variability in students' study strategies [49]. Factors beyond learning styles, such as motivation and teaching strategies, significantly impact learning outcomes, suggesting that effective preparation methods, including chunking, are critical for academic success [50]. Overall, the evidence indicates that Group B's increased engagement during finals week may be attributed to an exam format that complemented their study habits, particularly the use of chunking to navigate complex material effectively.

The literature illustrates a significant difference in student engagement and learning outcomes between chunked and unchunked assessments. Distributed practice—breaking material into smaller segments—yields better retention and comprehension than cramming [51]. Students who worked on split levels in a physics course achieved a mastery rate of about 70%, compared to only 30% for those facing the entire level at once, indicating that chunked assessments reduce unnecessary practice time and bolster student confidence (Gutmann et al., 2018). Similarly, increased segmentation in multimedia learning enhances recall and application, although excessive segmentation can lead to student dissatisfaction [52]. Shorter, chunked online modules result in higher completion rates and improved retention, showcasing the effectiveness of chunking in managing cognitive load [53]. Collectively, these studies indicate that chunked assessments foster greater student engagement and deeper cognitive processing compared to unchunked formats, which may overwhelm learners and hinder their success.

7.5 RELATIONSHIP OF ENGAGEMENT AND FINAL ASSESSMENT PERFORMANCE

Research consistently indicates a strong positive relationship between student engagement and academic performance, particularly in e-learning environments. Higher engagement levels in initial learning activities predict better performance in final assessments; however, this relationship is nuanced, as some students who excel early may underperform later, suggesting that factors beyond mere participation, such as individual learning strategies and prioritization of other modules, significantly impact outcomes [54]. Distinct groups of students based on engagement levels reveal that highly engaged students achieve significantly better scores compared to their less engaged counterparts, reinforcing the notion that engagement is critical for academic success but varies among students, indicating a need for tailored engagement strategies [55].

High engagement levels can lead to significant performance improvements, particularly for lower-competence readers, suggesting that engagement acts as a crucial motivating factor for those struggling academically [56]. The relationship between engagement metrics and performance has also been quantified, demonstrating that students with higher engagement consistently achieve better grades, underscoring the predictive value of engagement and emphasizing the importance of identifying and supporting less engaged students to improve academic outcomes [57].

A more complex view is provided by examining how engagement interacts with digital technology, revealing that low-performing students often struggle with distractions in technology-enhanced learning environments, which can detract from their academic success [58]. This highlights the necessity for educators to not only foster engagement but also equip students with self-regulation skills that enable them to manage distractions effectively. Together, these studies illustrate that while engagement is a vital determinant of academic performance, its effects are influenced by a myriad of factors, including the type of learning environment and individual student characteristics.

7.6 CHUNKED ASSESSMENT AS MORE ACCURATE REPRESENTATION OF STUDENT KNOWLEDGE

We posit that chunked assessments provide a more accurate representation of student knowledge than unchunked exams. The accuracy of measuring student performance is evaluated through scores, student engagement, and understanding of the material across different assessment formats. Our findings show that student knowledge levels before the final assessment were comparable across both cohorts, indicating that performance discrepancies arise from the assessment format. Specifically, chunked assessments led to higher scores and deeper understanding of the material. By dividing the final exam into manageable sections due on different days, students engaged more thoroughly, reducing testing fatigue and promoting retention.

In contrast, the unchunked format, requiring all sections to be completed in one sitting, resulted in lower scores and a less accurate representation of knowledge. This design caused increased testing fatigue, hindering students' ability to demonstrate their true understanding.

Chunked assessments also encourage self-regulated learning, allowing students to reflect on their performance and adjust their study strategies. This iterative process fosters better retention and a more reliable measure of student performance over time.

In conclusion, the implementation of chunked assessments enhances the accuracy of measuring student performance by yielding higher scores, promoting engagement, and providing a clearer representation of knowledge. This evidence supports the transition to chunked exams as an effective strategy for improving learning outcomes and assessments of student comprehension.

7.7 STUDENT COURSE ENGAGEMENT

The modified assessment structure's impact on student completion and performance on the final exam was quite pronounced. For Group A, the flexibility in choosing when to complete the exam led to a significant number of students postponing their submissions until after the deadline, as detailed in Figure 9. This procrastination was compounded by the exam structure, which required students to complete lengthy Interest and Self-Efficacy Surveys before they could address the technical content. Such a sequence likely contributed to mental fatigue, which, when coupled with the prerequisite of finishing earlier assessments, may have detrimentally affected their performance accuracy—reflected in the lower mean scores on the days leading up to and following the deadline.

Conversely, Group B experienced a more regimented exam schedule with explicit deadlines and a revised format that omitted the initial technical quizzes. This strategic adjustment was designed to minimize fatigue and optimize performance accuracy. By positioning the Interest and Self-Efficacy Surveys at the conclusion of the exam, students were able to concentrate on the technical material without prior cognitive load, thereby potentially enhancing their preparation and overall performance. As substantiated by hypothesis testing, Group B's final exam results were significantly better than those of Group A. This improvement is attributed to the structured exam format, which not only mitigated fatigue through well-defined study segments but also facilitated a more targeted review of the course material. The focused approach allowed students to build a deeper understanding and retention of the topics, which was evidently beneficial to their performance. The comparative analysis of the two groups underscores the efficacy of a structured assessment in promoting better academic outcomes, suggesting a reevaluation of unstructured exam formats in favor of more guided, segmented approaches to learning and evaluation.

7.8 FINALS WEEK COMPLETION RATE

The modified assessment structure's impact on student completion and performance on the final exam was quite pronounced. For Group A, the flexibility in choosing when to complete the exam led to a significant number of students postponing their submissions until after the deadline, as detailed in Figure 9. This procrastination was compounded by the exam structure, which required students to complete lengthy Interest and Self-Efficacy Surveys before they could address the technical content. Such a sequence likely contributed to mental fatigue, which, when coupled with the prerequisite of finishing earlier assessments, may have detrimentally affected their performance accuracy—reflected in the lower mean scores on the days leading up to and following the deadline.

Conversely, Group B experienced a more regimented exam schedule with explicit deadlines and a revised format that omitted the initial technical quizzes. This strategic adjustment was designed to minimize fatigue and optimize performance accuracy. By positioning the Interest and Self-Efficacy Surveys at the conclusion of the exam, students were able to concentrate on the technical material without prior cognitive load, thereby potentially enhancing their preparation and overall performance. As substantiated by hypothesis testing, Group B's final exam results were significantly better than those of Group A. This improvement is attributed to the structured exam format, which not only mitigated fatigue through well-defined study segments but also facilitated a more targeted review of the course material. The focused approach allowed students to build a deeper understanding and retention of the topics, which was evidently beneficial to their performance. The comparative analysis of the two groups underscores the efficacy of a structured assessment in promoting better academic outcomes, suggesting a reevaluation of unstructured exam formats in favor of more guided, segmented approaches to learning and evaluation.

8 CONCLUSION

In conclusion, our comprehensive analysis has highlighted critical insights into the academic performance and engagement levels of two distinct study groups within the course framework. Notably, the final exam scores and completion rates presented a stark contrast between the groups, with Group B outperforming Group A significantly. This difference in performance can be largely attributed to the assessment structure, which played a pivotal role in influencing study behaviors and exam preparedness.

The Interest and Self-Efficacy Surveys revealed nuanced motivational factors that potentially impacted the students' engagement trajectories. Group A exhibited high initial interest that fluctuated, possibly due to varying levels of self-efficacy. Conversely, Group B maintained a consistent sense of self-efficacy and engagement, which likely contributed to their sustained academic performance.

Despite similar technical quiz scores at the outset, indicating an equal footing in terms of knowledge and engagement, the two groups diverged in their final exam outcomes. This divergence underscores the significance of the assessment structure. Group A's flexible approach to exam scheduling and the sequential nature of their exam led to procrastination and possible cognitive fatigue. In contrast, Group B's structured assessment format, which allowed for focused and strategic preparation, proved to be more conducive to their learning style and ultimately resulted in higher performance.

These observations underscore the profound impact that assessment design can have on student outcomes. They highlight the importance of constructing assessments that not only evaluate student knowledge but also align with their study habits and preparation strategies to maximize performance and engagement.

Moving forward, our research raises several pertinent questions that warrant further investigation:

1. How does the performance on initial assessments influence students' engagement with the course material and their study patterns?
2. What is the relationship between early assessment performance and the likelihood of students revisiting the material for additional study?
3. Can regular feedback on the content of student performance, along with proactive strategies like explaining incorrect answers and offering encouragement, enhance student engagement and improve outcomes?
4. What are the effects of different cognitive-level questions on student performance, and how do students with diverse learning styles and demographics navigate their learning in an online course?

Addressing these questions could lead to a deeper understanding of educational assessment strategies and their impact on student learning. Through continued research in this area, we can refine our assessment methods to better cater to the diverse needs of students, thereby enhancing the overall educational experience and academic success.

APPROVAL DOCUMENTS

Institutional Review Board Approval Document

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Office of the Vice Chancellor for Research & Innovation

Office for the Protection of Research Subjects
1901 S. First St., Suite A, MC-685
Champaign, IL 61820

Notice of Approval: New Submission

December 13, 2023

Principal Investigator	Joshua Rovey
CC	Scott Nguyen
Protocol Title	*The Effect of Assessment Structure on Perceived Efficacy of a Rocketry Course*
Protocol Number	24497
Funding Source	Department of Defense, STEM: Expanding the Pipeline and Enhancing Education #HQ00342010040
Review Type	Expedited 6, 7
Status	Active
Risk Determination	No more than minimal risk
Approval Date	December 13, 2023
Expiration Date	December 12, 2024

This letter authorizes the use of human subjects in the above protocol. The University of Illinois at Urbana-Champaign Institutional Review Board (IRB) has reviewed and approved the research study as described.

The Principal Investigator of this study is responsible for:

- Conducting research in a manner consistent with the requirements of the University and federal regulations found at 45 CFR 46.
- Using the approved consent documents, with the footer, from this approved package.
- Requesting approval from the IRB prior to implementing modifications.
- Notifying OPRS of any problems involving human subjects, including unanticipated events, participant complaints, or protocol deviations.
- Notifying OPRS of the completion of the study.

DoD supported researchers must report the following within 30 days to the DoD human research protection officer:

1. When significant changes to the research protocol are approved by the IRB.
2. The results of the IRB continuing review.
3. Change of reviewing IRB.
4. When the University of Illinois Urbana-Champaign is notified by any Federal department, agency or national organization that any part of its HRPP is under investigation for cause involving a DoD-supported research protocol.

Department of Defense Approval Document

E02804.3a - OHRO Approval Memorandum (Proposal Number 21000256, Award Number HQ00342010040

1. The University of Illinois Urbana-Champaign (UIUC) Institutional Review Board (IRB) approved the above-referenced protocol application on 13 December 2023. The U.S. Army Medical Research and Development Command (USAMRDC), Office of Human and Animal Research Oversight (OHARO), Office of Human Research Oversight (OHRO) reviewed the protocol and found that it complies with applicable DoD, U.S. Army, and USAMRDC human subjects protection requirements.

2. The USAMRDC OHARO OHRO approves this no greater than minimal risk study for the enrollment of approximately 50 subjects.

3. The Principal Investigator must provide the following post-approval submissions to the OHRO via email to usarmy.detrick.medcom-usamrmc.other.mrmc-cr-documents@health.mil. **Failure to comply could result in suspension or termination of funding.** Send the following for OHRO review within the specified timelines:

a. **Prior to implementation of a substantive modification** – all documents related to substantive modifications to the research protocol and any modifications that could potentially increase risk to subjects. Substantive modifications include change in Principal Investigator, elimination or alteration of the consent process, change to the study population that has regulatory implications (e.g., adding children, adding active duty population, etc.), significant change in study design (i.e., would prompt additional scientific review), or a change in research procedures that could potentially increase risks to subjects.

b. **Prior to use of DoD funds for a new/additional performance site** – the site-specific protocol documents, IRB approval letter, study team members' qualifications documents.

c. **Upon change of the reviewing IRB** – IRB application/protocol and other documents approved by the new IRB, IRB approval letter.

d. **As soon as possible after receipt of re-approval from the IRB** – the progress report and a copy of the IRB continuing review approval letter. It appears that continuing review by the IRB is due no later than 12 December 2024.

e. **As soon as all documents become available** – the final study report submitted to the IRB, including a copy of any acknowledgement documentation and any supporting documents.

4. Promptly report the following study events via email to the OHRO by email to usarmy.detrick.medcom-usamrmc.other.hrpo@health.mil or by telephone (301-619-2165). Provide all supporting documentation to include the report to the IRB, IRB determination, corrective action plan, and any required follow-up.

a. All unanticipated problems involving risk to subjects or others.

b. Suspensions, clinical holds (voluntary or involuntary), or terminations of this research by the IRB, the institution, the sponsor, or regulatory agencies.

c. Any instances of serious or continuing noncompliance with the federal regulations or IRB requirements.

d. The knowledge of any pending compliance inspection/visit by the Food and Drug Administration (FDA), Office for Human Research Protections, or other government agency concerning this clinical investigation or research.

e. The issuance of inspection reports, FDA Form 483, warning letters, or actions taken by any government regulatory agencies.

f. Change in subject status when a previously enrolled human subject becomes a prisoner.

g. Note: Events or protocol reports received by the OHRO that do not meet reporting requirements identified within this memorandum will be included in the OHRO study file but will not be acknowledged.

5. **Please note:** The USAMRDC OHARO OHRO conducts site visits as part of its responsibility for compliance oversight. The study team must maintain accurate and complete study records in a secure and confidential manner, and make them available to representatives of the USAMRDC. Please note that the OHRO may contact the study team for additional information and documentation for the purpose of routine study monitoring at any time during award performance.

6. Do not construe this correspondence as approval for any contract or grant/cooperative agreement funding. Contact the appropriate contract/grants specialist or Contracting/Grants Officer regarding the expenditure of funds for your project.

7. The OHRO point of contact for this approval is Ms. Kristin Jones, Human Subjects Protection Scientist, at 301-619-7550/kristin.j.jones5.ctr@health.mil.

8. The OHRO point of contact for post-approval oversight is Ms. Jocelyn Rudolph, Human Subjects Protection Scientist, at 301-619-7550/jocelyn.s.rudolph.ctr@health.mil.

REFERENCES

[1] P. L. Ackerman and R. Kanfer, “Test Length and Cognitive Fatigue: An Empirical Examination of Effects on Performance and Test-Taker Reactions,” *J Exp Psychol Appl*, vol. 15, no. 2, pp. 163–181, Jun. 2009, doi: 10.1037/a0015719.

[2] D. J. Glaser and M. A. Insler, “The deleterious effects of fatigue on final exam performance,” *Econ Educ Rev*, vol. 90, p. 102305, Oct. 2022, doi: 10.1016/J.ECONEDUREV.2022.102305.

[3] P. L. Ackerman, R. Kanfer, S. W. Shapiro, S. Newton, and M. E. Beier, “Cognitive Fatigue During Testing: An Examination of Trait, Time-on-Task, and Strategy Influences,” *Hum Perform*, vol. 23, no. 5, pp. 381–402, Oct. 2010, doi: 10.1080/08959285.2010.517720.

[4] B. Humphries and D. Clark, “An examination of student preference for traditional didactic or chunking teaching strategies in an online learning environment,” *Research in Learning Technology*, vol. 29, pp. 1–12, Jan. 2021, doi: 10.25304/RLT.V29.2405.

[5] Center for Teaching & Learning, “How Do I Chunk Content to Increase Learning? : Center for Teaching & Learning : UMass Amherst.” Accessed: Nov. 22, 2024. [Online]. Available: https://www.umass.edu/ctl/how-do-i-chunk-content-increase-learning

[6] T. S. Faith, “Assessing Selected Teaching Techniques and Their Impact on Student Success in the Classroom,” *Teaching and Learning Excellence through Scholarship*, vol. 3, no. 1, Aug. 2023, doi: 10.52938/TALES.V3I1.2888.

[7] J. Domenech, D. Blazquez, E. de la Poza, and A. Muñoz-Miquel, “Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain,” *Educ Assess Eval Account*, vol. 27, no. 2, pp. 153–169, May 2015, doi: 10.1007/S11092-014-9208-Z.

[8] W. Kerdijk, J. Cohen-Schotanus, B. F. Mulder, F. L. H. Muntinghe, and R. A. Tio, “Cumulative versus end-of-course assessment: effects on self-study time and test performance,” *Med Educ*, vol. 49, no. 7, pp. 709–716, Jul. 2015, doi: 10.1111/MEDU.12756.

[9] D. Cecilio-Fernandes, M. Nagtegaal, G. Noordzij, and R. A. Tio, “Cumulative assessment: Does it improve students’ knowledge acquisition and retention?,” *Sci Med (Porto Alegre)*, vol. 28, no. 4, 2018, doi: 10.15448/1980-6108.2018.4.31880.

[10] R. A. Tio, M. E. Stegmann, J. Koerts, T. W. D. P. Van Os, and J. Cohen-Schotanus, “Weak self-directed learning skills hamper performance in cumulative assessment,” *Med Teach*, vol. 38, no. 4, pp. 421–423, Apr. 2016, doi: 10.3109/0142159X.2015.1132411.

[11] S. Ghosh, B. Brooks, D. Ranmuthugala, and M. Bowles, “Authentic Versus Traditional Assessment: An Empirical Study Investigating the Difference in Seafarer Students’ Academic Achievement,” *The Journal of Navigation*, vol. 73, no. 4, pp. 797–812, Jul. 2020, doi: 10.1017/S0373463319000894.

[12] T. Fawns and C. O’Shea, “Distributed Learning and Isolated Testing: Tensions in Traditional Assessment Practices,” *Proceedings of the International Conference on Networked Learning* , vol. 11, pp. 132–139, May 2018, doi: 10.54337/NLC.V11.8747.

[13] A. W. J. P. den Boer, P. P. J. L. Verkoeijen, and A. E. G. Heijltjes, “Comparing Formative and Summative Cumulative Assessment: Two Field Experiments in an Applied University Engineering Course,” *https://doi.org/10.1177/1475725720971946*, vol. 20, no. 1, pp. 128–143, Nov. 2020, doi: 10.1177/1475725720971946.

[14] K. Georgiou and I. Nikolaou, “Are applicants in favor of traditional or gamified assessment methods? Exploring applicant reactions towards a gamified selection method,” *Comput Human Behav*, vol. 109, p. 106356, Aug. 2020, doi: 10.1016/J.CHB.2020.106356.

[15] S. R. Sletten, “Rethinking Assessment: Replacing Traditional Exams with Paper Reviews,” *J Microbiol Biol Educ*, vol. 22, no. 2, Sep. 2021, doi: 10.1128/JMBE.00109-21/SUPPL_FILE/JMBE00109-21_SUPP_1_SEQ2.DOCX.

[16] N. Holmes, “Engaging with assessment: Increasing student engagement through continuous assessment,” *https://doi.org/10.1177/1469787417723230*, vol. 19, no. 1, pp. 23–34, Aug. 2017, doi: 10.1177/1469787417723230.

[17] M. Andrews, R. Brown, and L. Mesher, “Engaging students with assessment and feedback: improving assessment for learning with students as partners,” May 2018, Accessed: Nov. 22, 2024. [Online]. Available: https://ojs.cumbria.ac.uk/index.php/prhe/article/view/460

[18] U. Jayasinghe, A. Dharmaratne, and A. Atukorale, “Students’ performance evaluation in online education system Vs traditional education system,” *Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation, REV 2015*, pp. 131–135, Apr. 2015, doi: 10.1109/REV.2015.7087277.

[19] N. C. Cirit, “Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment.,” *Turkish Online Journal of Educational Technology - TOJET*, vol. 14, no. 3, pp. 9–19, Jul. 2015.

[20] C. T. McKevitt, “Engaging students with self-assessment and tutor feedback to improve performance and support assessment capacity,” *Journal of University Teaching and Learning Practice*, vol. 13, no. 1, Jan. 2016, doi: 10.53761/1.13.1.2.

[21] B. Huisman, N. Saab, J. van Driel, and P. van den Broek, "Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance," *Assess Eval High Educ*, vol. 43, no. 6, pp. 955–968, Aug. 2018, doi: 10.1080/02602938.2018.1424318.

[22] D. Vyas, J. Halilovic, M.-K. Kim, M. C. Ravnan, E. L. Rogan, and S. M. Galal, "Use of Cumulative Assessments in U.S. Schools and Colleges of Pharmacy," *Pharmacy 2015, Vol. 3, Pages 27-38*, vol. 3, no. 2, pp. 27–38, Jun. 2015, doi: 10.3390/PHARMACY3020027.

[23] V. Muniasamy, I. Magboul Ejalani, and M. Anandhavalli, "Predicting The Student?s Learning Outcome Based on Comparing the Assessment Methods in Diploma E-Commerce Course, Community College, King Khalid University, KSA," *International Journal of Web Technology*, vol. 004, no. 002, pp. 51–53, Dec. 2015, doi: 10.20894/IJWT.104.004.002.005.

[24] V. Kozlov, E. Alontseva, and A. Guryanov, "INFORMATION AND EDUCATIONAL ENVIRONMENT FOR MONITORING AND FORMING A CUMULATIVE ASSESSMENT OF A STUDENT'S ACADEMIC WORK IN A SEMESTER," *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference*, vol. 5, pp. 383–392, May 2019, doi: 10.17770/SIE2019VOL5.3828.

[25] H. H. Sievertsen, F. Gino, and M. Piovesan, "Cognitive fatigue influences students' performance on standardized tests," *Proc Natl Acad Sci U S A*, vol. 113, no. 10, pp. 2621–2624,Mar.2016,doi:10.1073/PNAS.1516947113/SUPPL_FILE/PNAS.1516947113.SAPP.PDF.

[26] A. Smith, "Cognitive Fatigue and the Wellbeing and Academic Attainment of University Students," *Journal of Education, Society and Behavioural Science*, vol. 24, no. 2, pp. 1–12, Feb. 2018, doi: 10.9734/JESBS/2018/39529.

[27] G. Borragán, H. Slama, M. Bartolomei, and P. Peigneux, “Cognitive fatigue: A Time-based Resource-sharing account,” *Cortex*, vol. 89, pp. 71–84, Apr. 2017, doi: 10.1016/J.CORTEX.2017.01.023.

[28] G. R. Wylie, B. Yao, J. Sandry, and J. DeLuca, “Using Signal Detection Theory to Better Understand Cognitive Fatigue,” *Front Psychol*, vol. 11, p. 579188, Jan. 2021, doi: 10.3389/FPSYG.2020.579188/BIBTEX.

[29] J. L. Jensen, D. A. Berry, and T. A. Kummer, “Investigating the Effects of Exam Length on Performance and Cognitive Fatigue,” *PLoS One*, vol. 8, no. 8, p. e70270, Aug. 2013, doi: 10.1371/JOURNAL.PONE.0070270.

[30] G. Borragan, H. Slama, A. Destrebecqz, and P. Peigneux, “Cognitive fatigue facilitates procedural sequence learning,” *Front Hum Neurosci*, vol. 10, no. MAR2016, p. 179603, Mar. 2016, doi: 10.3389/FNHUM.2016.00086/BIBTEX.

[31] Drexel University Disability Resources, “Chunking of Exams FAQ,” 2024. Accessed: Nov. 22,2024.[Online].Available:https://drexel.edu/disability-resources/~/media/Drexel/Treasurer-Group/Disability-Resources/Documents/Chunking-of-Exams.pdf

[32] M. Thalmann, A. S. Souza, and K. Oberauer, “How does chunking help working memory?,” *J Exp Psychol Learn Mem Cogn*, vol. 45, no. 1, pp. 37–55, Jan. 2019, doi: 10.1037/XLM0000578.

[33] T. Faith, “Assessing Selected Teaching Techniques and Their Impact on Student Success in the Classroom,” *Teaching and Learning Excellence through Scholarship*, vol. 3, no. 1, Aug. 2023, doi: 10.52938/TALES.V3I1.2888.

[34] J. R. Grewe, D. Kohler, M. C. Colver, T. N. Thurston, E. Wadsworth-Anderson, and M. Lewis, "Improving University Student Course Performance Using Reflective Exam Analysis: An Approach to Developing Academic Proficiency," *Scholarsh Teach Learn Psychol*, vol. 7, no. 4, pp. 253–262, 2021, doi: 10.1037/STL0000266.

[35] D. Grando, "What chunking of microbiology learning materials looks like today using technology and learning objects," *Microbiol Aust*, vol. 44, no. 3, pp. 127–130, Jul. 2023, doi: 10.1071/MA23037.

[36] D. J. Lees-Murdock *et al.*, "Assessing the Efficacy of Active Learning to Support Student Performance Across Undergraduate Programmes in Biomedical Science," *Br J Biomed Sci*, vol. 81, p. 12148, Mar. 2024, doi: 10.3389/BJBS.2024.12148/BIBTEX.

[37] S. Schweder and D. Raufelder, "Students' interest and self-efficacy and the impact of changing learning environments," *Contemp Educ Psychol*, vol. 70, Jul. 2022, doi: 10.1016/J.CEDPSYCH.2022.102082.

[38] L. K. Fryer, H. N. Bovee, and K. Nakao, "Self-efficacy latent growth trajectories' longitudinal links with achievement and interest: Both baseline and growth rate are important for interest outcomes," Nov. 2021, doi: 10.31219/OSF.IO/KHCDM.

[39] L. E. Hirshfield and D. Chachra, "Experience is not mastery: Unexpected interactions between project task choice and measures of academic confidence and self-efficacy in first-year engineering students," *International Journal of Engineering Education*, 2019.

[40] K. Talsma, K. Norris, and B. Schüz, "First-Year Students' Academic Self-Efficacy Calibration: Differences by Task Type, Domain Specificity, Student Achievement Level, and over Time.," *Student Success*, vol. 11, no. 2, pp. 109–121, 2020.

[41] M. de Fátima Goulão, “The Relationship between Self-Efficacy and Academic Achievement in Adults’ Learners.,” *Athens Journal of Education*, vol. 1, no. 3, pp. 237–246, Aug. 2014.

[42] P. J. Rottinghaus, L. M. Larson, and F. H. Borgen, “The relation of self-efficacy and interests: a meta-analysis of 60 samples,” *J Vocat Behav*, vol. 62, no. 2, pp. 221–236, Apr. 2003, doi: 10.1016/S0001-8791(02)00039-8.

[43] C. Entress and A. Wagner, “Beyond ‘Hitting the Books,’” *The Science Teacher*, vol. 081, no. 04, 2014, doi: 10.2505/4/TST14_081_04_27.

[44] A. Harris, S. Buglass, and G. Gous, “The impact of lecture chunking format on university student vigilance: Implications for classroom pedagogy,” *Journal of Pedagogical Sociology and Psychology*, vol. 3, no. 2, pp. 90–102, Oct. 2021, doi: 10.33902/JPSP.2021272429.

[45] T. T. Lineweaver, A. C. G. Hall, D. Hilycord, and S. E. Vitelli, “Introducing and Evaluating a ‘Study Smarter, Not Harder’ Study Tips Presentation Offered to Incoming Students at a Four-Year University.,” *Journal of the Scholarship of Teaching and Learning*, vol. 19, no. 2, pp. 16–46, Mar. 2019.

[46] I. A. E. Spanjers, T. Van Gog, and J. J. G. Van Merriënboer, “Segmentation of Worked Examples: Effects on Cognitive Load and Learning,” *Appl Cogn Psychol*, vol. 26, no. 3, pp. 352–358, May 2012, doi: 10.1002/ACP.1832.

[47] S. S. Stephenson, “Learning Styles Assessment Modalities Preferences Diagnostics (Lamp-D): A Framework Of Accounting Studentsâ€™ Preferred Learning Styles And Course Learning Assessments,” *The Accounting Educators’ Journal*, vol. 29, no. 1, Dec. 2019, Accessed:Nov.25,2024.[Online].Available:https://aejournal.com/ojs/index.php/aej/article/view/593

[48] I. Simonova and P. Poulova, “Assessment evaluation and preferences of engineering students,” *ITHET 2014 - 13th International Conference on Information Technology Based Higher Education and Training*, 2014, doi: 10.1109/ITHET.2014.7155675.

[49] T. Wilkinson, M. Boohan, and M. Stevenson, “Does learning style influence academic performance in different forms of assessment?,” *J Anat*, vol. 224, no. 3, pp. 304–308, Mar. 2014, doi: 10.1111/JOA.12126.

[50] E. N. Hidayati, A. Hidayat, and U. Ukit, “Relationship Between Learning Style With Student Learning Outcomes on Structure and Function of Plants,” *Pedagonal : Jurnal Ilmiah Pendidikan*, vol. 5, no. 2, pp. 94–105, Oct. 2021, doi: 10.33751/PEDAGONAL.V5I2.3794.

[51] M. Wiseheart, C. E. Küpper-Tetzel, T. Weston, A. S. N. Kim, I. V. Kapler, and V. Foot-Seymour, “Enhancing the Quality of Student Learning Using Distributed Practice,” *The Cambridge Handbook of Cognition and Education*, pp. 550–584, Jan. 2019, doi: 10.1017/9781108235631.023.

[52] P. E. Doolittle, L. H. Bryant, and J. R. Chittum, “Effects of degree of segmentation and learner disposition on multimedia learning,” *British Journal of Educational Technology*, vol. 46, no. 6, pp. 1333–1343, Nov. 2015, doi: 10.1111/BJET.12203.

[53] D. Méndez-Carbajo and S. A. Wolla, “Segmenting Educational Content: Long-Form vs. Short-Form Online Learning Modules,” *American Journal of Distance Education*, vol. 33, no. 2, pp. 108–119, Apr. 2019, doi: 10.1080/08923647.2019.1583514.

[54] B. Y. Rajabalee, M. I. Santally, and F. Rennie, “A study of the relationship between students’ engagement and their academic performances in an eLearning environment,” *E-Learning and Digital Media*, vol. 17, no. 1, pp. 1–20, Jan. 2020, doi:

10.1177/2042753019882567/ASSET/IMAGES/LARGE/10.1177_2042753019882567-FIG4.JPEG.

[55] F. A. Orji and J. Vassileva, "Modeling the Impact of Motivation Factors on Students' Study Strategies and Performance Using Machine Learning," *https://doi.org/10.1177/00472395231191139*, vol. 52, no. 2, pp. 274–296, Aug. 2023, doi: 10.1177/00472395231191139.

[56] G. C. Marchand and C. J. Furrer, "Formative, informative, and summative assessment: The relationship among curriculum-based measurement of reading, classroom engagement, and reading performance," *Psychol Sch*, vol. 51, no. 7, pp. 659–676, 2014, doi: 10.1002/PITS.21779.

[57] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Relationship between student engagement and performance in e-learning environment using association rules," *EDUNINE 2018 - 2nd IEEE World Engineering Education Conference: The Role of Professional Associations in Contemporaneous Engineer Careers, Proceedings*, Aug. 2018, doi: 10.1109/EDUNINE.2018.8451005.

[58] N. Bergdahl, J. Nouri, and U. Fors, "Disengagement, engagement and digital skills in technology-enhanced learning," *Educ Inf Technol (Dordr)*, vol. 25, no. 2, pp. 957–983, Mar. 2020, doi: 10.1007/S10639-019-09998-W.