Every Click You Make, Every Break You Take, We'll Be Watching You

David M. Pabst
Dept. of Mechanical Engineering and Materials Science
University of Pittsburgh
Pittsburgh, PA, USA
dmp114@pitt.edu

Lucas J. Zwastetzky
Dept. of Mechanical Engineering and Materials Science
University of Pittsburgh
Pittsburgh, PA, USA
ljz17@pitt.edu

Kristian A. Borysiak
Dept. of Mechanical Engineering and Materials Science
University of Pittsburgh
Pittsburgh, PA, USA
kab473@pitt.edu

Lee A. Dosse
Dept. of Mechanical Engineering and Materials Science
University of Pittsburgh
Pittsburgh, PA, USA
lad132@pitt.edu

Matthew M. Barry
Dept. of Mechanical Engineering and Materials Science
University of Pittsburgh
Pittsburgh, PA, USA
mmb49@pitt.edu

***Abstract*—This research-based full paper investigates Pre-recorded Lecture Video (PLV) viewership habits and elucidates correlations between practice and course performance. Using PLVs and Flipped Class Formats (FCFs) is increasingly prevalent in colleges and universities, driven by various motivating factors. Delivering content that was traditionally done didactically in a format that is both accessible and engaging has many benefits to not only the students but also instructors. For example, students can review PLVs multiple times at their convenience until they understand the material. From an instructional standpoint, PLVs and FCFs allow for more classroom interaction, whether through instructor-led examples or peer-to-peer learning activities. As with any new instructional technique, there are best practices that maximize both student performance and satisfaction, often guided by research. Previous studies have investigated both student use and interaction mechanisms of PLVs, specifically their level of viewership and completion of conceptual check-point questions that could be answered while or after watching assigned videos. Students who had substantial viewership and interacted with the PLVs by answering questions witnessed a statistically significant increase in their course grades compared to those who did not. However, viewership and interaction with a PLV do not tell the whole story. To this end, it is crucial to understand how students use PLVs, specifically, what habits they form while watching PLVs and how this behavior affects their performance in the course.**

During two years, 445 students enrolled in a sophomore-level Statics and Mechanics of Materials course were instructed using PLVs and an FCF. The students were required to view PLVs before class (on the University's approved Learning Management System, Canvas, using Panopto) and answer basic conceptual questions on Top Hat. Student viewership data (date and time watched, minutes delivered, etc.) was collected through Panopto. PLV viewership habits were ascertained: full-watching (e.g., watching and skipping, watching and rewinding), partial-watching (e.g., incomplete watching, skipping), and no viewership. Additionally, normalized viewership and average number of views per video were characterized. Correlations are drawn between viewership habits and final course grades. Students in the lower-performing group displayed poor habits with the PLVs (i.e., they had low normalized viewership and did not view many videos). In contrast, the middle- and higher-performing students displayed much better viewing habits (i.e., the opposite of low-performing students) and earned higher final course grades. It was also seen that the Top Hat cohort had greater upward mobility than their Traditional textbook counterparts.

***Index Terms*—Flipped class format, online lecture videos, engagement with digital media**

I. Introduction

The traditional education model has recently experienced a significant transformation, primarily driven by technological advancements and pedagogical innovations. One particularly favored approach that has gained traction in educational institutions worldwide is the Flipped Classroom Format (FCF). Traditionally, instructors would conduct didactic lectures during class time, resulting in a teacher-centric approach. However, the FCF "flips" this arrangement by moving the lecture component outside the classroom, allowing class time to be more student-centered. This approach enables students to engage with course content before class through Pre-recorded Lecture Videos (PLVs). This, in turn, creates time during class for questions and enhances interaction among students, the instructor, and teaching assistants [1]. Interaction is facilitated through various activities such as instructor-led example problems and collaborative in-class peer worksheets. It has been shown that implementing this newer pedagogy increases student performance [2]. In addition, exposure to this educational approach has improved student perceptions of the FCF, with students finding it superior to traditional teaching styles, which will be detailed in the following.

A study [3] conducted in 2020 surveyed undergraduate Civil Engineering students (n = 37) to gain insight into their FCF course content and learning activities; this was done through an anonymous online survey. The online survey consisted of Likert scale questions where the highest possible rating was indicated by "5−Completely Agree" and the lowest score indicated by "1−Completely Disagree". Over 95% of student respondents indicated they agreed or completely agreed with the course content being comprehensive, the course content being understandable, and the learning activities being suitable to understand the course topics. Following this survey, three focus groups (n = 6, n = 5, and n = 4) were conducted to gain further insight into how the students felt about the class while affirming the survey results. The interviews revealed that many students favored the FCF and suggested that they would like it applied to other courses. In another study conducted at the University of North Dakota, students who participated in an FCF were asked which class format they

preferred: the FCF or the traditional lecture. The results indicated that they preferred the FCF over the conventional lecture. Multiple studies have shown similar results, with students indicating they enjoy the application of the FCF to their courses [4]–[9].

FCFs have been shown to positively affect student performance in the classroom. A study [10] conducted recently compared student performance in a Mechanical Engineering course where one cohort was taught as a traditional didactic lecture (n = 99) and the other an FCF (n = 313). This study was conducted over multiple semesters. To assess the student's performance in the course, their three exam scores and final grades were compared between the two cohorts. A series of hierarchical multiple regression analyses were performed on their exams and final grades to establish if gender, ethnicity, or the classroom format affected the students' performance. It was shown that classroom format had a statistically significant effect on their first and second exam scores, where the FCF cohort was more likely to score higher than the traditional cohort. Similar results have been seen in both a Software Engineering course [11] and an Introductory Computer Design course [12], where again, the FCF cohorts outperformed the traditional lecture cohorts. The FCF has been shown to increase class performance, but interestingly enough, Dang et al. saw the most significant improvement occurred among the bottom two-thirds of their FCF course [13].

The flipped classroom model hinges on students engaging with lecture materials before class, effectively replacing the traditional in-class lecture. This typically involves providing PLVs for students to watch before class. Understanding how students interact with these videos is vital for gauging their performance in the course. In a study [14] conducted by Dazo et al. on an undergraduate Computer Science course, the effectiveness of the FCF was examined across three semesters between the fall of 2014 and the fall of 2015. The study tracked student interactions with PLVs and compared them to their course grades. Three key variables were devised to assess student engagement: videos viewed, content coverage, and punctuality. The 'videos viewed' variable measured the percentage of PLVs accessed out of the total provided for the course. 'Content coverage' offered insights into the percentage of each video watched by students. Lastly, 'punctuality' indicated how many hours students engaged with the videos before or after class, with positive numbers denoting viewing before class and negative numbers after. Initial findings from the fall 2014 semester revealed a significant lack of engagement with the videos. To address this, reflective questions were added to the PLVs in the spring 2015 semester. However, this did not substantially improve engagement. Consequently, midway through the semester, participation in discussions related to the videos was made mandatory. This participation requirement continued into the following semesters.

To investigate the relationship between the students' viewing habits and class performance, a Spearman correlation was conducted between the students' final course grades and each of the three variables described earlier. For the fall 2014 semester, all the variables positively correlated significantly with the final course grade. In the spring and fall 2015 semesters, only the 'punctuality' variable positively correlated significantly with the final course grade. It was believed that more correlations were present in the first semester because a large portion of the students were not engaging with the videos. Therefore, the smaller portion of students who adequately prepared for the course were likely to perform better than their counterparts. Though these metrics provide insights into how students' performance is potentially linked to their viewing habits, they do not thoroughly quantify how students engage with the PLVs themselves.

Given the extensive research demonstrating the positive impact of FCFs on student performance and the limited exploration into how high-achieving students utilize this format and their potential benefits for low-performing students, this study aims to investigate the correlation between students' behaviors when watching PLVs and their academic outcomes. Previous research [15] into how students engaged with PLVs and lecture video questions showed students who engaged with lecture video questions after watching the PLVs earned half a letter grade higher than those who did these tasks in reverse order; students with a higher normalized viewership earned almost a whole letter grade better in the course than students with lower viewership metrics; students who had high correctness to attempts ratios (i.e., earnestly tried to answer the lecture video questions) earned a whole letter grade higher than students who merely attempted to obtain participation points. Prior research has established a foundation for understanding how students interact with PLVs and how these behaviors might correlate with their performance in the course. However, these metrics typically reflect aggregated data from entire semester-long courses, which does not offer a nuanced understanding of students' specific viewing habits. The primary objective of this paper is to explore whether there are discernible differences in the PLV viewing habits of high-, medium–, and low-performing students.

II. Methodology

Table I provides basic demographic information about the population of students within this study, specifically their major and level of study. Notably, Mechanical Engineering and Bioengineering majors have the most significant percentages of students at 44% and 37%, respectively. The course is primarily taken by sophomore-level students, who make up 68.4% of the total population, followed by junior-level students at 25.8%. Within the sophomore cohort, Mechanical Engineering students are the largest group, accounting for 34.6% of the total, followed by Bioengineering students at 22.9%. The gender distribution shows approximately 68% of students were men, with women comprising the majority of the remainder and a small percentage identifying as "other."

The researchers examined the relationship between students' viewing habits and academic performance by analyzing data from an undergraduate Statics and Mechanics of Materials course over two years. This particular course was taught by three instructors and adhered to an FCF model. One of the researchers served as the primary instructor for all but two of the seven sections involved in the study. The students in all sections were tasked with viewing PLVs and completing corresponding video questions before attending class. The PLVs were available through the university's Learning Management System, Canvas, while the video questions were administered online through Top Hat. The Fall 2022 and Fall 2023 semesters had 66 and 70 PLVs, respectively, with a comprehensive set of 175 video questions aimed at assessing students' grasp of fundamental course material and their ability to recall information covered in the videos (reflecting the foundational level of Bloom's taxonomy [16]). Although more videos were added in the following semester, no additional video questions were created. Full course development and design details are provided in [17]. It is noted that of the seven sections, two used traditional textbooks, while the remaining five used an interactive, online textbook through Top Hat [18]. The conventional textbook sections were examined independently from those using the online interactive textbook.

Seven variables were created and observed to understand how student PLV viewing behaviors affected their performance in the course or if there were noticeable habit changes between the student

TABLE I
BASIC DEMOGRAPHICS OF STUDENT PARTICIPANTS IN THE STUDY, NAMELY MAJOR AND LEVEL WITHIN THEIR ACADEMIC CAREER.

	Major										
Level	Mechanical Engineering	Bioengineering	Materials Science Engineering	Industrial Engineering	Engineering Science	Civil Engineering	Chemical Engineering	Electrical Engineering	Duquesne 3+2 Program	Undeclared	Total
Freshman	0	0	0	0	0	0	0	0	1	0	1
Sophomore	154	102	17	13	1	0	0	0	0	19	306
Junior	35	56	10	4	6	1	0	1	0	2	115
Senior	6	5	0	8	2	1	1	0	0	0	23
Total	195	163	27	25	9	2	1	1	1	21	445

groups. The authors emphasize that this research is exploratory, with the selected variables intended to uncover specific PLV-viewing habits associated with high student performance. The students from each section were categorized into three groups based on their final course grades to be analyzed against the various variables to explain potential viewership habits. The first variable, "Average Normalized Viewership" ("Avg. Norm. View."), was calculated by dividing the total time a student spent watching each video by the video's duration. This process was repeated for every unique view of a lecture video, with the average computed over the total number of unique views for each student. A unique view was defined as any instance where a student watched a video at a different time, even if the same video was viewed more than once, such as on separate days. That is to say, if the student watched the video on Tuesday and then watched the same video again on Wednesday, that would be two unique views for that one lecture video.

The second variable "Full" represents instances where the length of a student's unique view matched the duration of the corresponding lecture video. It is important to note that this does not guarantee the student watched the entire video continuously. This will be discussed in greater detail later. The total number of "Full" views was summed to form the "Total Full" variable, which was subsequently analyzed. The third variable, "Total Partial," was defined by instances where a student's unique view duration was shorter than the full length of the lecture video. Each occurrence was recorded, and the totals were computed. Notably, the values for both "Total Full" and "Total Partial" can exceed the total number of videos, as students may watch videos multiple times.

When a student fast-forwards in any part of a unique view, this is noted as a "Skip." All the "Skips" for each unique view are then summed to form the variable denoted as "Total Skip." Whenever a student goes back in time while engaging with a lecture video, this is counted as a "Rewind" for a unique view. Again, these were all summed to form a variable adeptly named "Total Rewind." "Skips" and "Rewinds" can occur more than once for a unique view, as one can imagine a student may purposely be rewinding or skipping ahead in search of a specific instance within a lecture video. "Skips" and "Rewinds" can not occur independently. That is, they must be applied to a "Full" or "Partial" view. This is discussed in greater detail below with the aid of Fig. 1.

Figure 1 illustrates how these variables were mapped to students' viewership habits. The "Video Duration" column represents the time duration of a single video from start to end. Based on the viewing habits, the "Habits" show what the corresponding variable categories would look like. It is seen that the variables "Full" and "Partial" can be applied independently. A student may either watch the video in its entirety or stop before completion without rewinds or skips. However, the "Skip" and "Rewind" variables cannot occur alone; they must be associated with a "Full" or "Partial" view. For example, a "Full+Rewind" occurs when a student rewinds a portion of the video, and the total viewing time meets or exceeds the video's duration. Similarly, a "Full+Skip+Rewind" can also occur. A "Full+Skip" is impossible, as the total viewing time would not match the video's length. A "Partial+Skip" occurs when a student watches a video but skips ahead at any point, with the total viewing time being less than the video duration. The same applies to"Partial+Rewind," and "Partial+Rewind+Skip" behaviors, where each behavior is counted individually. An additional variable, "Did Not View" (DNV), not shown in Fig. 1, was used to record instances where students did not view a lecture video.

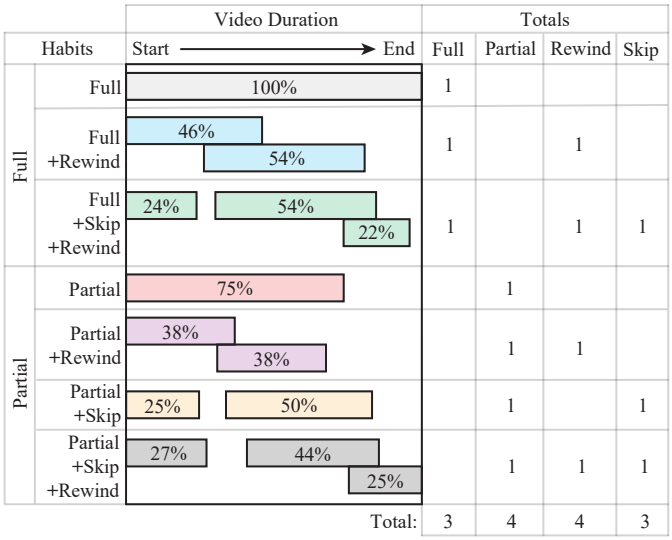

Fig. 1. Descriptions of possible viewing habits based on defined variables. The percentages are used as an example; 100% viewership would indicate a "Full" watch behavior, and anything less places the viewership behavior into the "Partial" category.

One-way between-subjects Analyses of Variances (ANOVAs) were run on the interactive online textbook cohort to elucidate differences between the groups and each of the variables discussed above. Before performing a one-way ANOVA, all assumptions were investigated to ensure the accuracy of the results. These assumptions, such as normality, homogeneity of variance, and significant outliers, were all assessed using IBM SPSS Statistics. To check for normality and homogeneity of variance, the Shapiro-Wilk and Levene's test for equality of variances was used, respectively. If a statistically significant difference was detected at a significance level of 0.05 among the groups, subsequent post-hoc analyses were performed to identify these variances; Tukey's Honestly Significant Difference (HSD) test was employed for pairwise comparisons.

The cohorts who used the traditional textbook were analyzed using a Kruskal-Wallis H-test to investigate if there were any statistically

significant differences between the groups. The Kruskal-Wallis *H*-test is a non-parametric test that is commonly used if the assumption of normality cannot be met [19], which was the case for all variables besides one, "Total DNV." This is noted in Tab. III.

III. Results and Discussion

The following sections present student behaviors related to PLVs for the interactive online and traditional textbooks. They also further investigate how a student's pre-existing GPA influenced their viewership habits and, ultimately, course grades.

A. Interactive Online Textbook Results

The analysis of the seven variables mentioned earlier has been conducted, and the findings will now be presented in the sequence in which they were introduced, starting with the cohort that used the interactive online textbook (denoted Top Hat). Shapiro-Wilk was used to determine normality before running a one-way ANOVA on these variables. All groups departed significantly from normality (p < 0.05) for all seven variables. However, since the sample sizes are roughly equal in size and substantially greater than 40, normality was assumed for all groups [20, 21].

A one-way ANOVA was performed between their final course grades and "Avg. Norm. View." to better understand what kind of viewership habits the different groups of students displayed and if they were significantly related to their final course grades. As mentioned earlier, normality was assumed due to the sample sizes, but Levene's test showed that the groups violated the homogeneity conditions. To account for this violation, a Welch's ANOVA was implemented instead [22]. It was determined that the groups differed significantly from one another, $F(2, 206.91) = 7.985$, $p < 0.001$. These results are summarized in Tab. II. Since the homogeneity of variances was violated, a Games-Howell post-hoc test was conducted for the pairwise comparisons [23]. This revealed statistically significant differences between the <80 group and both the $80-86.\bar{9}$ and $\geq$87 groups with p-values of 0.001 and 0.026, respectively. The post-hoc analysis revealed that the lowest performing tercile (<80) significantly differed from the other two terciles ($80-86.\bar{9}$ and $\geq$87), with p-values of 0.001 and 0.026.

Next, the researchers were interested in whether the "Total Full" watching habits impacted the students' performance in the course. Similar to earlier, normality was assumed, and significant differences between the variances were indicated. A Welch's ANOVA was used to examine if there were any statistically significant differences between the groups. The results indicated a significant difference, $F(2, 222.98) = 7.985$, $p = 0.002$. A Games-Howell post-hoc analysis revealed significant differences between the <80 group and both the $80-86.\bar{9}$ and $\geq$87 groups, with corresponding p-values of 0.001 and 0.026, respectively.

A one-way ANOVA was implemented to discern if there was a difference between the groups for the "Total Partial" variable. Beforehand, normality was again assumed, and the conditions for homogeneity were met based on the results of Levene's test (p = 0.055). The one-way ANOVA revealed a statistically significant difference between groups ($F(2,348) = 4.063$, $p = 0.018$). A pairwise comparison of means was performed to determine which groups were different from one another. Tukey's HSD showed the only difference that occurred was between the <80 and $\geq$87 groups ($p = 0.013$).

The following variables, "Total Skip" and "Total Rewind," did not satisfy the conditions for a one-way ANOVA as they both violated homogeneity. Therefore, Welch's ANOVA was used to determine if there were any differences between the groups for both variables. It was seen that the groups, when compared with the aforementioned variables, did differ significantly from one another. The results for the "Total Skip" and "Total Rewind" were $F(2, 219.12) = 8.820$, $p < 0.001$ and $F(2, 209.75) = 8.000$, $p < 0.001$, respectively. The variable mentioned above, "Total Skip," displayed statistically significant differences between the lower tercile and both the $80-86.\bar{9}$ ($p < 0.001$) and $\geq$87 ($p = 0.012$) terciles. Meanwhile, only the <80 and $80-86.\bar{9}$ groups differed ($p < 0.001$) from one another for the "Total Rewind" variable.

The last two variables satisfied all conditions for a one-way ANOVA, as normality was assumed based on the sample sizes. Levene's Test determined that the groups did not break homogeneity. The ANOVA showed a significant difference between the groups regarding the variables "Total DNV" and Average Number of Views ("Avg. No. Views"). The results were $F(2, 348) = 11.006$, $p < 0.001$ for the "Total DNV" variable and $F(2, 348) = 5.368$, $p = 0.005$ for the "Avg. No. Views" variable. The differences in groups for the "Total DNV" occurred between the <80, and both $80-86.\bar{9}$ and $\geq$87 groups, with a p-value of < 0.001 for each. The "Avg. No. Views" variable had a noticeable change between the lower and higher terciles with a p-value = 0.004, as Tukey's HSD post-hoc analysis indicated.

B. Interactive Online Textbook Discussion

As depicted in Fig. 2 a), all the mean plots of the seven variables have been normalized and laid over one another to help visualize not only one habit but how the numerous habits of the grouping of students are compared against their final course grade. This was done because looking at one habit individually does not encompass the totality of the students' habits, as they appear to be intertwined. The data shown in Fig. 2 a) aligns with what was discussed earlier, where it is noticeable that the lower tercile (<80) group of students had statistically significantly different viewing habits than the students grouped in the middle and higher performing terciles ($80-86.\bar{9}$ and $\geq$87).

Looking at the "Avg. Norm. View." in Fig. 3 a), there is about a 40% and 25% difference between the lower and middle terciles and the lower and higher terciles, respectively. It seems logical that as "Avg. Norm. View." Increased among the groups, and the performance increased within the course. One would expect that the more time you spent engaging with the videos, the better your understanding of the course material, and the higher your final grade would be. Interestingly, the middle tercile had about 25% higher "Avg. Norm. View." than the higher performing tercile, yet they did not perform as well. This difference was not found to be statistically significant, however looking at this one variable alone might not give enough detail about the viewing habits of students.

The normalized mean plot in Fig.3 a) shows a linearly increasing trend from the lower to higher performing terciles. The most significant increase occurred from the lower to middle tercile with an increase of about 35% in "Total Full" views. Then, an approximate increase of 10% from the middle to higher tercile. The significant increases seen between the lower and both the middle and higher terciles align with the post-hoc test results discussed above. It appears that the higher-performing students had more "Full" views than the lower-performing students, and this is shown to be statistically significant. It seems that watching the video in "Full" may positively affect the student's performance.

The slope for "Total Partial" appears to be consistent across the three terciles, with a 20% increase from the lower to middle tercile and another 20% increase from the middle to higher tercile, approximately. Tukey's HSD revealed that the only statistically significant

TABLE II
DESCRIPTIVE STATISTICS AND COMPARISON OF MEANS OF TOP HAT DATA SET. GROUP $\geq$87, $n = 124$; 80$-$86.$\bar{9}$, $n = 105$; <80, $n = 122$.

		Group Statistics		Levene's Test for Equality of Variances	One-Way ANOVA		Welch's ANOVA		Shapiro-Wilk Test of Normality
Variable	Group	Mean	σ	Significance	F	Significance	F	Significance	Significance
Avg. Norm. View.	$\geq$87	1.81	2.14	< 0.001	-	-	7.985	< 0.001	< 0.001
	80$-$86.$\bar{9}$	2.49	3.44						< 0.001
	<80	1.17	1.67						< 0.001
Total Full	$\geq$87	10.52	12.09	0.011	-	-	6.447	0.002	< 0.001
	80$-$86.$\bar{9}$	9.65	11.97						< 0.001
	<80	5.95	9.54						< 0.001
Total Partial	$\geq$87	54.81	61.12	0.055	4.063	0.018	-	-	< 0.001
	80$-$86.$\bar{9}$	46.06	59.43						< 0.001
	<80	34.73	44.92						< 0.001
Total Skip	$\geq$87	157.62	155.10	0.002	-	-	8.820	< 0.001	< 0.001
	80$-$86.$\bar{9}$	186.31	180.10						< 0.001
	<80	105.44	126.75						< 0.001
Total Rewind	$\geq$87	26.15	33.85	< 0.001	-	-	8.000	< 0.001	< 0.001
	80$-$86.$\bar{9}$	37.47	46.12						< 0.001
	<80	18.10	25.21						< 0.001
Total DNV	$\geq$87	37.34	20.56	0.527	11.006	< 0.001	-	-	< 0.001
	80$-$86.$\bar{9}$	37.29	19.74						< 0.001
	<80	47.77	19.495						< 0.001
Avg. No. Views	$\geq$87	0.95	0.94	0.061	3.985	0.005	-	-	< 0.001
	80$-$86.$\bar{9}$	0.81	0.93						< 0.001
	<80	0.59	0.71						< 0.001

difference is between the lower and higher terciles, where the higher tercile had about 40% more "Partial Views" than the lower tercile. This leads to the idea that the more students interact with the videos, whether a "Full" or "Partial" view, the better they perform.

"Total Skip" and "Total Rewind" exhibit trends similar to "Avg. Norm. View." All three metrics show a substantial increase from the lower to the middle terciles, followed by a smaller decrease from the middle to the higher tercile. Significant variations were again seen between the lower and middle terciles for both variables, with a significant change being seen between the middle and higher terciles for the "Total Skip" variable. The middle and higher performing terciles spend much less time skipping and rewinding than the lower terciles. This, paired with the earlier observation of "Avg. Norm. View.," shows that, in general, the students in the lower tercile are not interacting with the PLVs nearly as much.

The "Avg. No. Views" mean plot is nearly identical to the "Total Partial" mean plot, which appears to have a linear increase with a constant slope. The increase from the lower to middle tercile and middle to higher tercile are approximately 15% and 20%, respectively. The "Total DNV" mean plot is the only one that does not replicate any of the earlier patterns mentioned. This initially decreases about 20% from the lower to middle tercile, then is almost invariant from the middle to higher tercile.

Overall, the pattern appears that students in the middle and higher terciles interact more with the videos than students in the lower terciles. This can be seen statistically through multiple post-hoc analyses, where the majority of differences were between the lower terciles and both the middle and higher terciles. Looking at the mean plots helps visualize these differences. It would appear that the lack of interaction with the videos or bad viewing habits has led to these students' lower final grades in the course.

C. Traditional Textbook Results

To observe what kind of habits different level-performing students have with the PLVs, the researchers again compared their final course grades to the seven variables previously described. However, for the cohorts using the traditional textbook, the distribution of the groups departed significantly from normality for all but one variable. The sample sizes of the groups were not large enough to assume normality. Therefore, a non-parametric Kruskal-Wallis H-Test [19] was used to identify if there were any statistically significant differences between the groups. The subsequent results are in Tab. III.

The three groups were compared against all but one variable, "Total DNV," using a Kruskal-Wallis H-Test. This showed a statistically significant difference between groups for the six variables as the p-values were below 0.05, as seen in Tab. III. To ascertain where the differences between the groups occurred, a Dunn's test [24] was used to conduct a pairwise analysis. In addition, a Bonferroni correction was implemented because multiple hypotheses were being tested, which increases the probability of observing a significant difference by chance. This means the possibility of obtaining a type I error (i.e., a false positive) rises. The Bonferroni correction adjusts for this by controlling the overall error rate, reducing the likelihood of false positives [25]. The outcome of these tests will be described in more detail below.

The "Avg. Norm. View." variable showed significant differences between the <78 tercile, and both the 78$-$84.$\bar{9}$ and $\geq$85 terciles, with adjusted p-values of 0.002 and 0.017, respectively. This same pattern was observed in multiple other variables: "Total Full," "Total Partial," "Total Rewind," and "Avg. No. Views." This pattern was not reflected in the "Total Skip" variable, where the only difference occurred between the lower tercile (<78) and the middle tercile (78$-$84.$\bar{9}$) with an adjusted p-value of 0.008.

Lastly, the groups were compared using Welch's ANOVA with the remaining variable, "Total DNV." This variable followed a normal distribution but violated the assumption of homogeneity of variances. The ANOVA revealed significant differences between the groups, $F(2,\ 59.093) = 11.098$, $p < 0.001$. A Games-Howell post-hoc test was subsequently conducted to identify where these differences occurred. Significant differences, $p < 0.001$, were once again seen

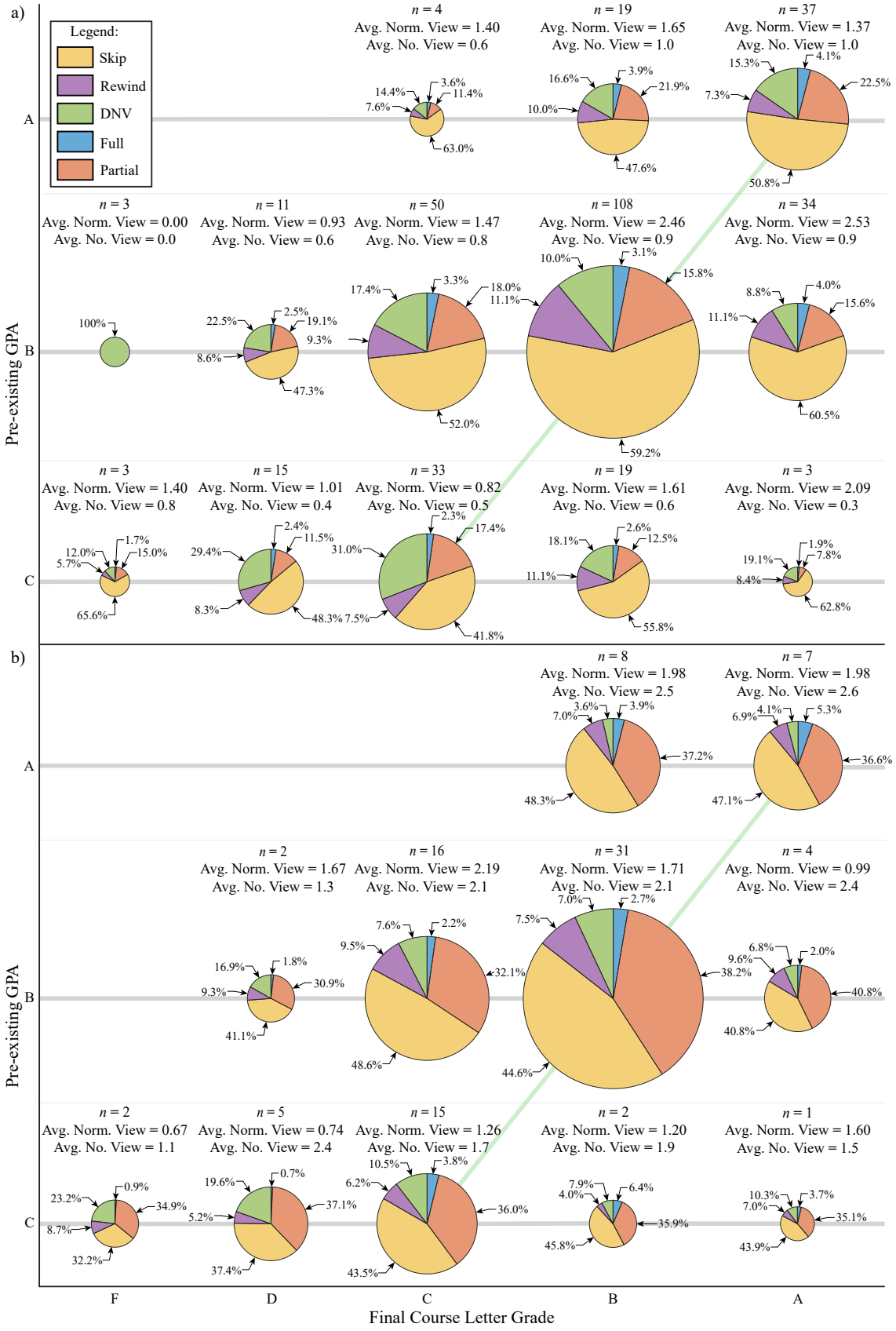

Fig. 2. Cohort metrics (population size n, average normalized viewership, the average number of views, and behavior: Skip, Rewind, DNV, Full and Partial) for Final Course Grade (A, B, C, D, and F) versus pre-existing GPA (A, B, and C) for students using the a) Top Hat and b) Traditional textbooks.

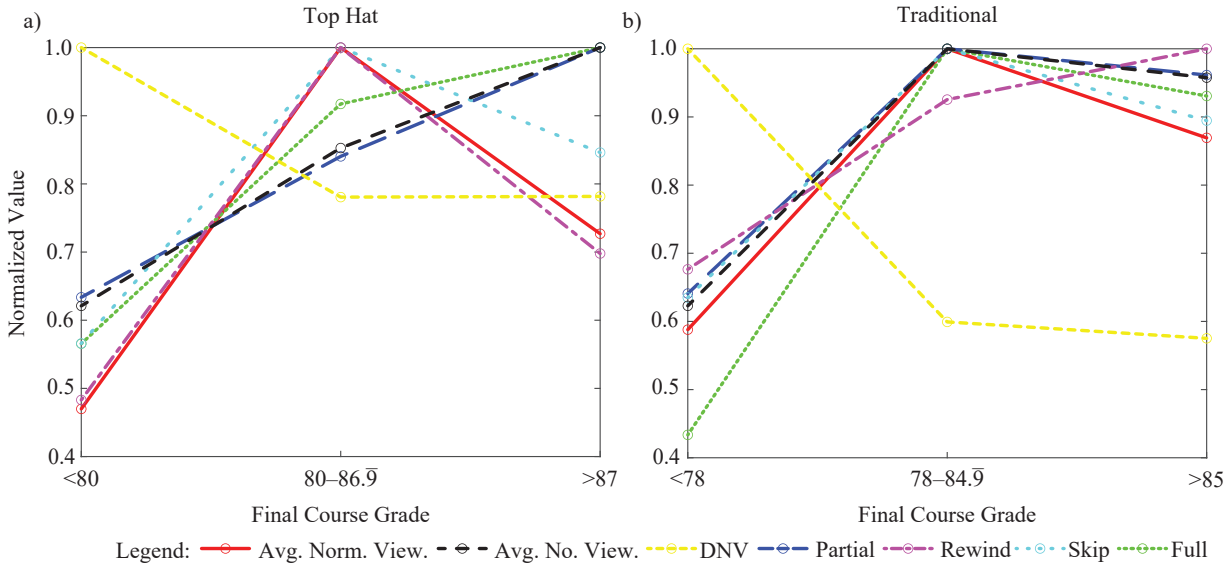

Fig. 3. Normalized Mean Plots of Variables for both the Traditional and Top Hat Textbook Cohorts

TABLE III
DESCRIPTIVE STATISTICS AND COMPARISON OF MEANS OF TRADITIONAL TEXT DATA SET. GROUP $\geq$85, $n = 29$; 78$-$84.$\bar{9}$, $n = 36$; <78, $n = 30$.

		Group Statistics		Levene's Test for Equality of Variances	Kruskal-Wallis H-Test		Welch's ANOVA		Shapiro-Wilk Test of Normality
Variable	Group	Mean	σ	Significance	H	Asymptotic Sig.	F	Significance	Significance
Avg. Norm. View.	$\geq$85 78$-$84.$\bar{9}$ <78	1.73 1.99 1.17	0.98 1.29 1.32	-	13.202	0.001	-	-	< 0.001 < 0.001 0.043
Total Full	$\geq$85 78$-$84.$\bar{9}$ <78	13.03 14.00 6.07	13.16 10.96 8.15	-	12.483	0.002	-	-	< 0.001 0.003 0.001
Total Partial	$\geq$85 78$-$84.$\bar{9}$ <78	143.81 148.81 95.37	66.60 57.17 77.87	-	11.563	0.003	-	-	0.028 0.031 0.017
Total Skip	$\geq$85 78$-$84.$\bar{9}$ <78	170.59 190.69 121.10	82.24 100.97 97.52	-	9.849	0.007	-	-	0.009 0.003 0.007
Total Rewind	$\geq$85 78$-$84.$\bar{9}$ <78	31.00 28.69 20.97	20.81 18.59 25.94	-	10.635	0.005	-	-	< 0.001 < 0.001 < 0.001
Total DNV	$\geq$85 78$-$84.$\bar{9}$ <78	22.45 23.39 39.03	12.34 13.14 16.50	0.046	-	-	11.098	< 0.001	0.159 0.390 0.058
Avg. No. Views	$\geq$85 78$-$84.$\bar{9}$ <78	2.26 2.34 1.48	0.96 0.87 1.16	-	13.802	0.001	-	-	0.049 0.011 0.004

between the <78 tercile, and both the 78$-$84.$\bar{9}$ and $\geq$85 terciles.

D. Traditional Textbook Discussion

The traditional mean plots in Fig. 3 b) reveal a pattern similar to what was observed in the Top Hat cohort. Most variables show a substantial increase (30-55%) from the lower to middle tercile, followed by a slight decrease (5-15%) towards the higher tercile. This pattern is evident in the "Avg. Norm. View.," "Avg. No. Views," "Total Partial," "Total Skip," and "Total Full" variables. This trend supports the idea that lower-performing students interact less with the PLVs than middle and high tercile students. A Dunn's pairwise comparison confirmed that the differences between the lower tercile and both the middle and higher terciles were statistically significant.

The "Total Rewind" variable exhibited a slightly different trend. Instead of peaking at the middle tercile, it increased by about 20% from the lower to middle tercile and continued to rise by another 10% from the middle to higher tercile. The increase from the lower to the higher terciles was statistically significant, further indicating that lower-performing students interact less with the videos than their peers. The "Total DNV" variable reinforces this finding.

For the "Total DNV" variable in the Traditional textbook cohort, the pattern mirrored that of the Top Hat textbook. There was a substantial decrease (around 40%) in the videos not viewed from the lower to middle tercile, which leveled off in the higher tercile. This supports the idea that lower-performing students' viewing habits negatively impact their final course grades. However, the broad grade

ranges within each tercile (with the lower tercile covering grades from C+ to F) suggest that a more granular classification might provide better insights.

E. Pre-Existing GPA Discussion

To better understand students' viewing habits and how they relate to their pre-existing and final course grades, Figs. 2 a) and b) were created. The y-axis is the student's pre-existing GPA when entering the course, and the x-axis is the student's final grade in the course. The size of the pie charts is determined by the number of students that fall into a specific location in the table. Note that the relative sizes between the two cohorts are different due to the Traditional textbook cohort having very small sample sizes for some of the populations. These pie charts are then divided into normalized percentages of their viewing habits, with the colors referencing specific viewing habits denoted by the legend. The sample size, "Avg. Norm. View.," and "Avg. No. Views" are listed, from top to bottom, above the pie charts for the various distinct outcomes. This was done for both the Top Hat and Traditional textbook groups. It's important to note that no statistical analysis was conducted on these figures, and any comments are strictly observational.

Comparing the habits of students who entered the course with an A and left with an A and those who entered the course with an A and left with a B, it is seen that their habits appear nearly identical for both the Top Hat and Traditional sections. This same pattern holds for students who entered the course with a B and left with an A and those who entered with a B and left with a B. This suggests that students with a pre-existing GPA of an A have similar habits, yet their performances differed by a letter grade. This situation is comparable to that of pre-existing B students. Unfortunately, there is still insufficient information to determine what might be causing this drop or increase in performance, although the habits are nearly identical. However, it suggests that students with GPAs corresponding to an A or B have similar habits that potentially enable them to perform higher than students entering with lower GPAs.

When comparing the Top Hat textbook students against the Traditional students, there appear to be some differences. First, students in the Traditional textbook cohort have a higher average number of views per video than the Top Hat groups. The "DNV" percentages are also higher in the Top Hat groups. This indicates that the students in the Traditional textbook group watched the videos more often than those who used the Top Hat textbook. One possibility is that the students in the Top Hat section were incentivized to use their book because the embedded questions in their book counted towards their participation grade. The students using the Traditional book, although instructed to read the textbook, had no incentive to use their book. Studies have shown that about 40% of students tend not to use their textbooks [17, 26, 27]. If the Traditional textbook cohort did not use their book as often as the other cohort, they may have relied more on the videos to further their understanding.

Looking at the students who entered with an A in the Traditional cohort, more than half left with a grade other than an A. Most students who entered with a B tended to leave with the same score (about 58%), with only about 8% leaving with an A and about 34% leaving with a lower score. Students who entered with a C also tended to leave with the same grade (approximately 60%), while about 28% received a lower grade and 12% received a higher grade.

Observing the Top Hat group, a higher percentage of students who entered with an A left with an A (62% compared to 53% in the Traditional cohort). Similarly, most students who entered with a B also left with a B (about 52%), but a more significant percentage improved their grades to an A (31% compared to 8% in the Traditional textbook cohort). This trend continued with the C students; a more significant percentage improved their grades (about 30% compared to 12% in the Traditional textbook cohort), 45% left with the same score, and only 24% left with a lower score. Students in the Top Hat cohort seemed more likely to leave the course with the same or an improved score than their Traditional cohort counterparts. That is to say, when viewing the green diagonal line or neutral line (those who entered and exited with the same GPA/course grade), there appears to be more upward grade mobility with students who used the Top Hat textbook than the Traditional textbook.

IV. Conclusion

The goal of this study was to see if students' viewing habits had an impact on their performance within a Statics and Mechanics of Materials I course. Seven variables were created using the quantitative data provided by Panopto to assess how the students interacted with the videos. To understand if any of the habits were statistically significant, multiple analyses were performed to include, but not limited to, ANOVAs, non-parametric tests, and pairwise comparisons. All seven variables proved to have a statistically significant effect on the students' final grades for both the Top Hat and Traditional textbook sections. Multiple post-hoc analyses were done to locate where the differences between the groups were occurring.

It was shown that significant differences in the Top Hat cohort appeared between the lower tercile (<80) and both the middle (80-86.$\bar{9}$) and higher ($\geq$87) terciles for the "Avg. Norm. View." (p = 0.001, p = 0.026), "Total Full (p = 0.031, p = 0.003)," "Total Skip (p < 0.001, p = 0.022)," and the "Total DNV (p < 0.001, p < 0.001)" variables. The "Total Partial" and "Avg. No. Views" saw differences between the lower and higher terciles with p-values of 0.013 and 0.004, respectively. The last variable, "Total Rewind," only saw differences between the lower and middle terciles with a p-value of < 0.001.

All variables, besides "Total Skip," saw significant differences between the lower (<78) and both the middle (78-84.$\bar{9}$) and higher ($\geq$85) terciles. The variables and their respective p-values are as follows: "Avg. Norm. View." (p = 0.002, p = 0.017), "Total Full (p = 0.002, p = 0.048)," "Total Partial (p = 0.004, p = 0.028)," "Total Rewind (p = 0.012, p = 0.016)," "Avg. No. Views" (p = 0.001, p = 0.013), and "Total DNV ($p \geq 0.001$, $p \geq 0.003$)." The final variable, "Total Skip," had a significant difference between the lower and higher tercile with a $p-$value of 0.008.

Though we recognized some differences between the lower terciles and the middle and higher terciles, we still did not gain much insight into the differences in habits between the middle and higher tercile students. Without corroborating the video-watching behaviors with students' interactions with their books, lecture video questions, or homework questions, we may be missing information that differentiates the higher-performing students from the middle-performing students. The data showed that students' habits range drastically in the middle and higher performing terciles, which can be seen based on the standard deviations provided in both Tabs. II and III. That is, we see students in the middle performing terciles who display similar habits to those in lower and higher terciles and vice versa. Yet, why do they perform differently from one another? We only have a small piece of information regarding what may influence student success.

In the future, researchers aim to investigate the timing of students' engagement with the PLVs, both before and after class. Previous studies have suggested that this may significantly impact students' performance in the course. Developing more robust variables to

understand students' viewing habits better is also essential. Instead of tracking individual skips and rewinds, we could categorize views as "Partial+Skip+Rewind." Another idea is to correlate the times students engage with the PLVs with their other course assignments, which may indicate that students are using the videos to enhance their understanding of the material alongside their coursework. In the future, it would be beneficial to monitor students' viewing habits for each lecture video and compare this data with their exam scores and final course grades. This analysis could provide insight into how students' viewership habits change over time and how such changes potentially impact their course performance. This future work should be applied to the Top Hat and Traditional textbook cohorts.

During this research, it was seen that the Top Hat cohort appeared to have a lower "Avg. Norm. View." values than the Traditional textbook cohort. It was noted in another study [28] that students would typically use the method they preferred when completing the course assignments in the Top Hat sections. Some students preferred using the PLVs to study the course material over the Top Hat textbook and, therefore, would only watch the PLVs, whereas some students felt the opposite; they would only use the Top Hat textbook and ignore the PLVs. This may be the cause for the change in "Avg. Norm. View.," but it may also be because the Top Hat group was incentivized to use the book over the PLVs.

REFERENCES

[1] J. Bergmann et al., "Flip your classroom: Reach every student in every class every day," Eugene, Or.: International Society for Technology in Education, 2012.

[2] R. M. Clark et al., "Flipping engineering courses: A school wide initiative," *American Society for Engineering Education: Advances in Engineering Education*, vol. 5, no. 3, pp. 1–39, 2016.

[3] H. Kayaduman, "Student interactions in a flipped classroom-based undergraduate engineering statistics course," *Computer applications in engineering education*, vol. 29, no. 4, pp.969–978, 2021.

[4] M. Cavalli et al., "Comparison of student performance and perceptions across multiple course delivery modes," in 2014 ASEE Annual Conference & Exposition, pp. 24–300, 2014.

[5] D. Battaglia et al., "How flipping your first-year digital circuits course positively affects student perceptions and learning," *The International journal of engineering education*, vol. 31, no. 4, pp. 1126–1138, 2015.

[6] G. Mason et al., "Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course," *IEEE transactions on education*, vol. 56, no. 4, pp. 430–435, 2013.

[7] P. Kiat et al., "The flipped classroom experience," in 2014 IEEE 27th Conference on Software Engineering Education and Training, pp. 39–43, 2014.

[8] D. Jean-Pierre et al., "Flipping the controls classroom around a MOOC," in 2014 American Control Conference, pp. 2557–2562, 2014.

[9] G. B. Johnson, "Student perceptions of the Flipped Classroom," T, University of British Columbia, 2013.

[10] H. Cho et al., "Active learning through the flipped classroom in mechanical engineering: improving students' perception of learning and performance," *International Journal of STEM Education*, vol. 8, pp. 1–13, 2021.

[11] L. Gren, "A flipped classroom approach to teaching empirical software engineering," *IEEE Transactions on Education*, vol. 63, no. 3, pp.155-163, 2020.

[12] A. Amresh et al., "Evaluating the effectiveness of flipped classrooms for teaching CS1," in 2013 IEEE Frontiers in Education Conference (FIE), pp. 733–735, 2013.

[13] Q.-V. Dang et al., "Bringing In-Class Online - A Hybrid Solution," in Fourth Interdisciplinary Engineering Design Education Conference, pp. 12–17, 2014.

[14] S. Dazo et al., "An empirical analysis of video viewing behaviors in flipped CS1 courses," *ACM Inroads*, vol. 7, no. 4, pp. 99–105, 2016.

[15] D. M. Pabst et al., "Student Engagement with Pre-Recorded Lecture Videos in a Flipped-Class Format," in *2023 IEEE Frontiers in Education Conference* (FIE), pp. 1–5, 2023.

[16] D. Krathwohl et al., "Merlin C. Wittrock and the revision of Bloom's taxonomy," *Educational psychologist*, vol. 45, no. 1, pp. 64–65, 2010.

[17] D. M. Pabst et al., "A Comparative Study on Student Performance using Traditional and Interactive Textbooks,"; in American Society of Engineering Education (ASEE) 2023 Annual Conference and Exposition, 2023.

[18] M. Barry et al., *Statics and Mechanics of Materials: An Example-based Approach*, Top Hat, ISBN: 978-1-77412-230-3, 2020.

[19] Laerd Statistics, "Kruskal-Wallis H Test in SPSS Statistics | Procedure, output and interpretation of the output using a relevant example.," Laerd.com, 2014. [Online]. Available: https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php. [Accessed: May 21, 2023].

[20] A. Elliott et al., *Statistical analysis quick reference guidebook: With SPSS examples*, Sage, 2007.

[21] J. Pallant, *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*, McGraw-hill education (UK), 2020.

[22] H. Liu, "Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance," Virginia Commonwealth University, 2015.

[23] "One-Way ANOVA Post Hoc Tests," www.ibm.com. [Online]. Available: https://www.ibm.com/docs/en/spss-statistics/SaaS?topic=anova-one-way-post-hoc-tests. [Accessed: May 21, 2023]

[24] O. J. Dunn, "Multiple Comparisons Among Means," *Journal of the American Statistical Association*, Vol. 56, no. 293, pp. 52–64, 1961.

[25] Y. Benjamini et al., "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, no. 1, pp. 289–300, 1995.

[26] K. Baier et al., "College students' textbook reading, or not," *American reading forum annual yearbook*, vol. 31, no. 31, pp. 385–402, 2011.

[27] T. Berry et al., "An exploratory analysis of textbook usage and study habits: Misperceptions and barriers to success," *College teaching*, vol. 59, no. 1, pp. 31–39, 2010.

[28] L. Dosse et al., "Beyond the Bind and Between the Lines: Focus Group Insights Into an Interactive-Online Textbook," in *2024 IEEE Frontiers in Education Conference* (FIE), 2024.