ABSTRACT

INVESTIGATION INTO THE USE OF A COLLABORATIVE E-BOOK READER AMONG INTRODUCTORY PHYSICS STUDENTS

by Patrick James Carroll

The reading and study habits of students is of great interest to educators and those who study education. Additionally, e-book usage is becoming more prevalent in the classroom. We have conducted a study on the usage and integration of a social annotation e-book reader, Perusall, in introductory physics classes. Perusall allows annotating and commenting in a social environment on an assigned reading from a text. These annotations are based on predefined settings and proprietary algorithms. We have implemented Perusall in introductory level, calculus-based physics courses for two academic years. The classes met three times a week and each class included a reading assignment due before the next class. We obtained data that included time information, the number of annotations made, and data on how the class as a whole used the textbook on a page-by-page basis, in terms of number of page views and time on the page. Along with this data we also have information on how the students’ final grades. Lastly, we have surveyed students on their experience using Perusall and collected demographic data. Here we examine the results of this two-year study on using Perusall in these classes and the students involved.
INVESTIGATION INTO THE USE OF A COLLABORATIVE E-BOOK READER AMONG INTRODUCTORY PHYSICS STUDENTS

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Science
Department of Physics
by
Patrick James Carroll
Miami University
Oxford, Ohio
2019

Advisor: Dr. Jennifer Blue
Reader: Dr. Stephen Alexander
Reader: Dr. Karthik Vishwanath

©2019 Patrick James Carroll
This Thesis titled

INVESTIGATION INTO THE USE OF A COLLABORATIVE E-BOOK READER AMONG INTRODUCTORY PHYSICS STUDENTS

by

Patrick James Carroll

has been approved for publication by

The College of Arts and Science

and

Department of Physics

____________________________________________________

Dr. Jennifer Blue

____________________________________________________

Dr. Stephen Alexander

____________________________________________________

Dr. Karthik Vishwanath
# TABLE OF CONTENTS

List of Figures ................................................................................................................ iv  
Dedication ......................................................................................................................... vi 
Acknowledgements .......................................................................................................... vii  
Introduction ...................................................................................................................... 1  
Lit Review .......................................................................................................................... 3  
Methods ............................................................................................................................. 6  
  Method of Instruction ....................................................................................................... 6  
  Perusall Configuration and Assignments ......................................................................... 7  
  Perusall Data Collection ................................................................................................. 8  
  Cautionary Note ............................................................................................................. 9  
  Survey ............................................................................................................................. 9  
  Exploratory Data Analysis ............................................................................................. 10  
  Viewing Time, Active Time and Number of Annotations ............................................... 11  
  Institutional Review Board ............................................................................................ 12  
Data ................................................................................................................................... 13  
  Survey Responses .......................................................................................................... 13  
  Initial Statistical Evaluation ........................................................................................... 14  
  Mean Viewing Time and Mean Active Time ................................................................ 25  
  Mean Active Time and Mean Number of Annotations ................................................... 30  
  Mean Active Time and Final Grade .............................................................................. 37  
  Number of Annotations and Final Grade ..................................................................... 47  
  Modification of Perusall Settings ................................................................................... 54  
  Page View Statistics ...................................................................................................... 55  
Discussion ......................................................................................................................... 59  
  General Results ............................................................................................................. 59  
  Demographic Results .................................................................................................... 61  
  Page View Results ........................................................................................................ 62  
  Implications for Instruction ............................................................................................ 63  
  Future Work ................................................................................................................... 65  
Bibliography ...................................................................................................................... 67  
Appendix A: Perusall Survey ............................................................................................ 69  
Appendix B: IRB Application ............................................................................................ 73
LIST OF FIGURES

Figure 1: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in this study..................................................14

Figure 2: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in academic year ’17 - ’18. .........................15

Figure 3: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in academic year ’18 - ’19. .........................15

Figure 4: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in PHY 191. ..............................................16

Figure 5: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in PHY 192. ..............................................16

Figure 6: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as female. .........................17

Figure 7: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as male. .........................17

Figure 8: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as international. .....................18

Figure 9: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as domestic. .........................18

Figure 10: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for freshman students. .............................................19

Figure 11: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for sophomore students. .................................19

Figure 12: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for junior students. ..........................................20

Figure 13: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for senior students. ...........................................20

Figure 14: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in STEM.............................................21

Figure 15: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students outside of STEM........................................21

Figure 16: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in health or kinesiology. .........................22

Figure 17: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in biology or microbiology. .......................22

Figure 18: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in the trailer sections. .........................23

Figure 19: The impact of mean viewing time on mean active time for all students. .......................25

Figure 20: The impact of mean viewing time on mean active time for students in the trailer section. ................................................................................................................26

Figure 21: The impact of mean viewing time on mean active time for academic year. ...................27

Figure 22: The impact of mean viewing time on mean active time for PHY 191 and PHY 192. ....27
Figure 23: The impact of mean viewing time on mean active time for students who identify as male or female. ................................................................. 28
Figure 24: The impact of mean viewing time on mean active time for domestic and international students. .............................................................................................................. 28
Figure 25: The impact of mean viewing time on mean active time for freshman and sophomores. ................................................................................................................................... 29
Figure 26: The impact of mean viewing time on mean active time for juniors and seniors. ....... 29
Figure 27: The impact of mean active time on mean number of annotations time for all students. ........................................................................................................................................ 30
Figure 28: The impact of mean active time on mean number of annotations time for academic year ‘17-‘18. ................................................................................................................. 31
Figure 29: The impact of mean active time on mean number of annotations time for academic year ‘18-‘19................................................................................................................... 31
Figure 30: The impact of mean active time on mean number of annotations time for PHY 191 and PHY 192. .................................................................................................................. 32
Figure 31: The impact of mean active time on mean number of annotations time for the trailer sections. ................................................................................................................... 33
Figure 32: The impact of mean active time on mean number of annotations time by semester. 34
Figure 33: The impact of mean active time on the mean number of annotations time by international status. .......................................................................................................................... 35
Figure 34: The impact of mean active time on the mean number of annotations time by international status. .......................................................................................................................... 36
Figure 35: The impact of mean active time on final grade for all students.............................. 37
Figure 36: The impact of mean active time on final grade for all students.............................. 38
Figure 37: The impact of mean active time on final grade for all students with lower bound. .... 39
Figure 38: The impact of mean active time on final grade for fit to the lower trendline............. 40
Figure 39: The impact of mean active time on final grade by academic year.......................... 41
Figure 40: The impact of mean active time on final grade for PHY 191 and PHY 192............... 42
Figure 41: The impact of mean active time on final grade for the trailer sections. ..................... 43
Figure 42: The impact of mean active time on final grade by gender. ........................................ 44
Figure 43: The impact of mean active time on final grade by international status. .................. 45
Figure 44: The impact of mean active time on final grade by class. ........................................ 46
Figure 45: The impact of mean number of annotations on final grade for all students............... 47
Figure 46: The impact of mean number of annotations on final grade by academic year......... 48
Figure 47: The impact of mean number of annotations on final grade for PHY 191 and PHY 192. ........................................................................................................................................ 49
Figure 48: The impact of mean number of annotations on final grade for trailer sections. ....... 50
Figure 49: The impact of mean number of annotations on final grade by gender. ..................... 51
Figure 50: The impact of mean number of annotations on final grade by international status. ... 52
Figure 51: The impact of mean number of annotations on final grade by class. ....................... 53
Figure 52: The number of times each page was viewed by each class by page number. .......... 55
Figure 53: The average amount of time each page was viewed by each class by page number. 56
Figure 54: The average amount of time each page was viewed by how many times the page was viewed by each class. .................................................................................................. 57
DEDICATION

To my mother, Cindy: Thank you for supporting me through all of this. I really couldn’t have done it without you.

To Megan: Thank you for being there for me and constantly cheering me on.

To my boys, Corbin and Gavin: I know this has been as hard for you both, but you both mean the world to me.

I love you all.
ACKNOWLEDGEMENTS

To my adviser, Dr. Jennifer Blue: Thank you for allowing me to run free and encouraging me to explore my research interests.

To Dr. Stephen Alexander, Dr. Christopher Beer, Dr. Bogdan Leu, and Patrick Boyle: Thank you for letting me use your classes as my research playground and working with me through all of this.

To Dr. Karthik Vishwanath: Thank you for introducing me to Perusall and talking with me through the whole process.
INTRODUCTION

Textbooks have been a part of higher education curriculums and learning for centuries. Today, textbooks are central to the construction, preparation, and implementation of modern physics classes. Students are expected to read and utilize their textbook as the central repository of information and knowledge pertaining to their class, in conjunction with lectures and labs. Additionally, the textbook is expected to be the resource students draw on to participate in discussions with their instructor and other students, both in and out of class. To facilitate the use of textbooks and these ongoing conversations within the modern college class, various software products and e-book readers have been developed to manage textbooks and assignments and encourage collaboration among students and educators.

Much of the research on textbooks and the benefits and habits of students regarding reading is contentious and, at times, pessimistic. It is, however, generally accepted by most in academia that reading one’s textbook is mandatory, if not beneficial. It is also a common belief, especially in student-led, upside-down pedagogies, that the more students interact with one another while discussing physics, the better they perform. Perusall is a social collaboration e-book reader that uses a proprietary grading system on annotations made by students in their digital textbook as part of their reading assignments. For two years, the Department of Physics at Miami University has implemented Perusall in one of their course offerings to engage both of these beliefs.

This thesis studies and quantifies the implementation of Perusall in an introductory level, calculus-based physics course. This study was conducted over a two-year period, for academic years ’17 - ’18 and ’18 - ’19, in which students completed a two-semester, comprehensive physics course, where both semesters used the same textbook. Each semester consisted of several sections of either PHY 191 or PHY 192 being taught by three different instructors, or trailer sections consisting of the opposite material, PHY 192 instead of PHY 191, for example, taught by a fourth instructor. The students were given a selection of online readings to complete each week and asked to make a specific number of annotations in order to receive full credit. At the conclusion of each course, all information on these annotations, grades, and the students’ behavior within Perusall was collected.

Various data was collected directly from Perusall, such as how many annotations students made, how long students spent in their textbook, how much students interacted within their textbook such as clicking and scrolling, and their grades both for both the assignments and the individual annotations. Data regarding the use of individual pages in terms of how many views it received, and the average time spent viewing that page, was available on a per class basis. The final grades of the students were also collected at the conclusion of each of the semesters. Additionally, a survey was conducted in order to collect demographic data and gauge students’ familiarity with Perusall, or similar technology, and determine their attitudes towards Perusall and its use.
Two major goals exist for this study of Perusall use in these courses: 1) to create a profile of how students complete their assignments and make use of their textbook and 2) to inform instruction both in Perusall use and general textbook use. This study should inform instructors as to how to best implement Perusall, its features, and the textbook as a whole to benefit students. At the same time, it should set the groundwork for future studies and inform potential areas for further exploring, both qualitative and quantitative. Exploratory data analysis (EDA) has been used extensively to these ends, employing a variety of statistical techniques and graphing to look at these complex sets of data.
Physics Education Research (PER) studies how individuals learn physics and the various ways we can improve the methods used in physics education. This is an extremely broad field which includes specialized areas in assessment, behavior, cognition, problem solving, societal issues, student and teach characteristics, classroom management, curriculum development, learning environment, and (of course) pedagogy. The methods used and desired outcomes for research like this can vary widely. This can range from developing and refining standardized testing to examining social and identity dynamics, such as gender, age, or origin. It can include evaluations of the language or content used in a course or even the technology or labs available in the classroom.

While much of what is done in PER is related to psychology and statistics, a deep understanding of physics is needed to be an effective researcher. Not only must one have an understanding of the fundamentals of a concept in order to be an effective instructor, but also have consideration for how these concepts fit into a larger picture, what impact these ideas and techniques have on other materials and courses, and an idea of what the current big topics are within physics to help lay the groundwork for it. Of course, a passion for the subject and enthusiasm for why it should be learned is helpful as well. In the broadest sense, PER has tried to answer questions such as “Why is physics so hard to learn?” and “What’s the difference between experts and novices?”, but at its core it wants to know “Are there more effective ways to teach?”.

Less than a century ago, education was commonly viewed as an art and not a science and often still is. Even the very first article of the very first issue of the American Journal of Physics included this statement by Floyd K. Richtmyer: “Teaching, I say, is an art, and not a science.” Through the 1970’s and 1980’s PER grew and research focused on trying to evaluate the disconnect related to specific concepts or the difference between instruction material and preconceived notions. In 1992, Hestenes, Swackhammer, and Wells published the Force Concept Inventory (FCI), which is a standardized way to measure a students’ understanding of basic concepts in Newtonian physics. This standardized test illustrated that students did not always have the conceptual understanding we think they do, even if they perform well on quantitative problems, and led to new fields of study. We still use the FCI in many of our classes today. In fact, 2018 was the first year since 2014 we have not given the FCI to PHY 191 students.

By the 2000’s, one area of great interest was in peer focused instruction and interactive classrooms. These methods focus less on the traditional lecture style of education and, instead, try to create collaborative and active learning environments wherein the students are encouraged to work together, investigate and explore the topics in an interactive capacity, and become part of the teaching process as much as the learning process. One example of this is the Student-Centered Active Learning Environment with Upside-down Pedagogies (SCALE-UP) which was developed at North Carolina State University and currently in use at Miami University in the
PHY 161/2 and PHY 191/2 classes. This style of learning integrates the traditional, but separate, lecture and lab combination into one collective learning experience.

A staggeringly large number of questions can still be formed within PER, but beyond the traditional questions of pedagogy and students’ understanding of the material, two areas of note will be big topics going forward. First, the impact of technology and the internet on the classroom and identity studies is rapidly evolving but has had limited research conducted on its impact. Additionally, some academic approaches have been slow to adopt these changes, favoring the more traditional lecture and reading/assignment combinations. Entire classrooms are now online and, one day, physics may be asked to do the same. But, beyond that, the ease of use and accessibility when it comes to programming abilities, open source products and texts, and augmented learning environments all leave much to be evaluated in the future of PER. Second, the area of identity studies, which has been a topic of study within science education research for several years but is newer to PER specifically. This includes gender studies, queer studies, disability and veterans’ issues, and cultural and socioeconomic studies. This seeks to determine the impact a person’s identity and current status in life has on their ability to learn physics.

Our PER group focuses on several areas, the four broad categories being: working with teachers, studying students, reforming instruction, and promoting equity. This includes summer institutes for practicing teachers, implementing Perusall (described below) into the PHY190’s classes to augment and study students’ reading habits, studying SCALE-UP, working with the engineering and psychology departments to study student behaviors and beliefs in the 190’s, and gender studies within the physics learning environment. We have been involved in other areas as well, such as reforming or updating labs at the 100 and 200 level, investigating inquiry-based science teaching, and evaluating incivility in the college classroom.

Textbooks have been a part of higher education curriculums and learning since the founding of the first universities. Today, textbooks are central to the construction, preparation, and implementation of modern physics classes. Students are expected to read and utilize their textbook as the central repository of information and knowledge pertaining to their class, in conjunction with lectures and labs. Additionally, the textbook is expected to be the resource students draw on to participate in discussions with their instructor and other students, both in and out of class. To facilitate the use of textbooks and these ongoing conversations within the modern college class, various software products and e-book readers have been developed to manage textbooks and assignments and encourage collaboration among students and educators.

While most physics students purchase a textbook, some literature reports as little as 20%-30% of students read their textbook and 60% of those read after lecture. One study in particular found that in STEM or humanities, regardless of seeking a two-year or four-year degree, only between 20%-40% of students completed their readings. Various strategies can be implemented to increase reading times of students. In some cases, implementing a pre-reading assignment, such as a reading quiz, has been shown to get up to 80% of physics students to complete their readings. Even still, students who claim to have done their reading do not always show proficiency in the subject matter afterwards, with one study citing only 55% of
students who claimed to have read their assignment could show even basic comprehension of the subject.\textsuperscript{14} It has been suggested that one possible reason for this lack of reading by students is that many students, of various majors, do not think the textbook is critical to their studies and do not read ahead in the material.\textsuperscript{15} How students use their textbook and what effect reading their textbook has on their course performance can be difficult to determine due to the variety of textbooks used, different types of reading assignments, lecture styles, and grading methods. However, it is generally accepted by instructors that reading is beneficial to students not only in terms of grades, but also with respect to their stress and feelings towards the subject matter studied.

Many new technologies and web-based platforms exist to assist instructors in getting information to students and to help those students make the best use of that information. This includes greater use of e-books, learning management systems, and other virtual learning environments. An example of these emerging technologies is social collaboration and online annotation environments where students engage with one another online through a common media, often their textbook. There is evidence to suggest that quality participation by students using this type of learning environment, specifically in a flipped or upside-down pedagogy, improves overall course performance by students.\textsuperscript{16} It has also been suggested that use of these types of systems increase positive and decrease negative emotions related to learning the course material.\textsuperscript{17} Lastly, the use of online annotation tools may also increase the interactivity and social engagement of students throughout the course.\textsuperscript{18} A study was also conducted by Perusall LLC in conjunction with Harvard University that suggested that Perusall, when adopted into an upside-down pedagogy, had over 90\% success rate in getting students to engage with their reading and an overall increase in test scores.\textsuperscript{19} However, this particular class structure contained no formal lecture and relied solely on the text for the classes instruction and was compared to a control group that was given similar reading assignment, only without the social annotation tools.

In this study, we have looked at the use of an e-book reader that features social collaboration and annotation tools, Perusall, to determine how students interact with their textbook and one another when given a reading assignment, as well as to determine if such an application can improve pre-class reading participation. This online application, designed at Harvard, uses behavioral and educational science with data analytics to track the progress of students through instructor defined reading assignments. Establishing the initial settings and reading assignments is the extent of the instructor’s involvement, after which engagement through collaborative annotations is student driven and grading is done automatically by Perusall. We investigated the level of participation on reading assignments by each student, examined the effects of participation by students on their final grade for the course, and conducted a survey to determine how students felt about using Perusall.
METHODS

As discussed, while textbooks have been studied in various ways, no one method exists as the gold standard for evaluating the impact of textbooks on students. Similarly, how textbooks are used by students is also a challenge to put defined metrics to for a number of reasons. For this study, an EDA approach has been used in order to evaluate a large volume and variety of data. EDA is a commonly used analysis approach which seeks to characterize a data set, beyond just the traditional hypothesis testing and statistical modeling. EDA is the main method used in this study to attempt to define what metrics, if any, we can obtain from the data regarding our methods of implementing Perusall in our classrooms, given our methods of instruction in the Department of Physics at Miami University. This involves understanding the construction of and methods used in our classrooms, the type and demographics of students taking the course, how Perusall itself functions and is setup, how the assignments have been presented, and the data we can retrieve from Perusall. Given these factors, some methods were refined semester-to-semester while others were unable to be changed or remained unchanged by choice, in order to avoid introducing yet more variables.

Method of Instruction

This study evaluates the use of Perusall in a freshman-level introductory physics class for science and engineering majors. This is a calculus-based physics course which used “Physics for Scientists and Engineers: A Strategic Approach with Modern Physics (4th Edition)” by Randall D. Knight. The content of this textbook was divided between two courses, PHY 191 and PHY 192, which were taught over the fall and spring semesters of academic years ’17 – ’18 and ’18 – ’19. PHY 191 covers classical mechanics and quantum physics while PHY 192 covers thermal physics, electromagnetism, and relativity. Each section contained up to 99 students, though most had fewer. The classes made use of the flipped classroom instructional method known as SCALE-UP. This method includes integrated lab and lecture, with student led and centered content, as well as online homework assignments using Mastering Physics, three mid-term exams, and one final exam. Students also received a participation grade during class, via Learning Catalytics. Students were seated three to a table in class and these seats were chosen for them, which determined the groups they worked in for the semester. These seatings were left up to the instructor to select, however most were done randomly.

These courses were taught by four different professors, three for the main sections and one for the trailer sections. Exam grades and final grades for each student were collected at the end of each semester. For most sections, Perusall assignments were worth approximately 5% of the student’s final grades. However, in the trailer section, the sections of PHY 192 taught in the fall and PHY 191 in the spring, had the student’s Perusall grade as 10% of the student’s final grade. Concern over how Perusall scores impact on final grade were expressed to the students
was discussed, but ultimately left up to the individual instructor to decide how this was presented. Each class varied in which final numerical grades translated to final letter grades and, in the end, what was considered passing in the class. However, in most cases, the grading scale was more generous than the traditional $90 = A$, $80 = B$, $70 = C$, etc., grading scale. As a result, a mean final grade of 70 was closer to a B, on average, than the traditional grade. This is significant to keep in mind when discussing what was passing or minimum requirements.

Perusall Configuration and Assignments

Implementing a system like Perusall allows instructors to assign readings where students are randomly placed into discussion groups and receive points for annotating the assigned section of the text. While the actual system used to grade these annotations is proprietary in nature, in general it assigns a weighted score to words that correspond to the text and compares the student’s comments using that information. The developer’s claim is that two out of three independent graders give the same score as Perusall. These grades are given on an annotation-by-annotation basis and averages the top comments made by a student; the exact number of annotations to be averaged is set by the instructor. The default grading scale for annotation grades is: 0 for deficient annotations, 1 for annotations that need improvement, and 2 for annotations that meet expectations. For a better understanding a scoring guide for Perusall can be found on their website. This was the scoring method used for these courses. Perusall is able to penalize students’ grades on an assignment based on how well distributed their comments are throughout that assignment. This can range from 0% to 100% penalty if the student falls under a threshold of commenting on at least 20% of the assignment. A 10% penalty was used for all courses presented here. The students’ annotations were scored immediately and recorded in their personal, online gradebook, allowing the student to return to their assignment and add or change the content as needed to receive an improved score.

Each Perusall course was setup separately for each class, meaning there was no interactions between the various classes, with one exception being the two trailer sections of PHY 191 in the Spring ‘19 semester, which combined both classes. When the students enroll in the Perusall course they are randomly assigned to a group of 20 students. Annotations made by a student can only be seen and responded to by members of their group. This is to ensure that each assignment has a significant number of annotations for students to interact with without overwhelming the student, as it has been suggested by previous research that this negatively impacts students grades. Perusall is also capable of notifying students, called a “nudge”, when an assignment is do, more annotations are needed, or someone has replied to that student’s annotation. Instructors are also provided with a “confusion report” which shows them the most interacted with pages of the text and the top questions asked by the students in their annotations, which this type of Just-in-Time-Teaching has been suggested to be beneficial to students and instructors. This is meant to be used to augment and assist in the instructor’s lesson planning, not to encourage instructors to monitor the flow of conversation within the text or to get
involved. In fact, instructors were specifically asked not to use Perusall themselves. This is to prevent the instructor from unintentionally ending conversations between students by being seen as the authority and subject matter expert.

Prereading assignments were given before each class period, in order for the students to familiarize themselves with the material beforehand. These assignments varied in length depending on the chapter and material covered. In the first academic year, students were asked to make a fixed number of annotations on each assignment, ten annotations, of which the top five were graded. In the subsequent year the students were asked to make a comment on each page of the assignment with the best scoring comments equal to half of the number of pages in the assignment being graded. The length of the assignments varied widely, as needed for that day’s lecture, but most ranged between four and ten pages, with the first assignment being closer to 30 pages. While all the instructors covered the same material at the same time, the actual content of the assignments themselves varied by instructor, with some including or excluding specific pages of the textbook. The PHY 191 classes also included an extra credit assignment in which the students were given one very long reading assignment that contained five chapters, needing fewer annotations, and having significantly longer time to complete it.

Perusall Data Collection

Various data was collected from the analytics tool provided by Perusall which fall into two broad categories: class data and student data. The class data included the average viewing time by the class as a whole for each page of the textbook and the number of times each page was viewed by the entire class. Unfortunately, this data was not available on a per student basis with the free version of Perusall’s analytics that we were making use of for these courses. Student data included viewing time, active time, number of annotations made, the grades for each assignment, and the scores for each individual annotation. The viewing time is the total time the student has their textbook open, regardless of activity level. This does not include time spent with the browser minimized or Perusall in a background tab or browser. The user is automatically logged out after 15 minutes of inactivity, even if Perusall is currently being viewed on the screen. Active time is defined as every time the user moves their mouse, interacts with their touch screen, or types on their keyboard and the two-minute period following that activity. Lastly, while Perusall differentiates between initial comments placed in the textbook, statement responses to those comments, and questions asked in the book or in the comments, we have opted to only consider whether a comment was made or not, regardless of the type.

Once the data was collected from Perusall and the individual instructors, MATLAB was used to clean, sort, and evaluate all the data. Much of this process was labor intensive early in our study given the amount of data available, the number of permutations possible, and the EDA nature of our approach. This involved correcting time formatting and artifacts from Perusall.
sorting students by section and demographic, all statistical evaluations and graphing, and alternative data exploration methods. As much of this process as possible was automated within MATLAB, including routine statistical evaluations and graphing, so that it would not have to be manually repeated at the end of each semester. However, the class data from Perusall was unavailable in a format that could easily be read into MATLAB and so had to be manually entered each time. This impacted the speed at which we could evaluate some forms of data in this study and future works.

Cautionary Note

Unfortunately, we have no control over the more proprietary aspects of Perusall and how it operates. This includes having access to its grading algorithm. While we know the general manner in which it calculates its scores and the quality control goals set forth by its creators, there are a few artifacts we cannot avoid. One artifact is that, on occasions due to differing browser configurations, times are counted differently for active time and viewing time. This is corrected to the best of our ability, usually by defaulting the viewing time to the active time, in the event that the active time is greater than the viewing time. This issue impacted less than 1% of students. The other artifact that needs to be addressed is a loophole in the grading algorithm that some students discovered and used to game the Perusall scoring system. If a student made a significantly long annotation it was automatically scored for full credit, regardless of the content of that annotation. Instructors and Graduate Assistants were warned about this issue and told to scan the comments each class period to weed out any that may have gotten credit for non-physics related topics. It was decided that if a topic was physics related, even if it got its score from using this loophole, it would receive the credit. We relied on this and students flagging inappropriate annotations to combat this issue, but only a limited qualitative check was conducted to ensure this issue did not bias the data too significantly.

Survey

Surveys were conducted midway through each semester, except for the Spring ’19 semester. These surveys included questions designed to determine students’ history with similar style of software, ease difficulty of use, general attitude towards Perusall or expectations related to it, and demographic data for the students. These surveys were generally conducted around the second midterm exam, so that the students had had significant time to acclimate to and form opinions about Perusall. These surveys were voluntary, and a large portion of the student population did not complete the surveys, however a statistically significant portion did complete the survey. The demographic data collected included gender, status as a domestic or international student, academic year, and category of major. The categories of majors were STEM, biology/health, non-STEM, and other. This survey has been included in Appendix A.
Exploratory Data Analysis

EDA techniques were used to evaluate the large collection of data available for possible trends or usable outcomes. This included examining means and standard deviations for each of our variables, analysis of variance (ANOVA) to compare across classes and demographics, general fitting techniques to attempt to determine correlations between elements of the data and using computational techniques to try to evaluate other trends in the data. Initially, means and standard deviations were calculated for each student’s viewing time, active time, number of annotations, and Perusall assignment grades. Sum totals for viewing time, active time, and number of annotations was also found. These were done for every demographic, by class type (PHY 191 and PHY 192), and by semester. Boxplots were made for each of these for ease of assessment and comparison, as well as to visually examine the outliers and quartiles of the respective data. These mean values for each variable pertaining to a student were plotted against one another and the students’ final grades in order to begin assessing their impact. It is important to note that when working with large data sets of varied data, simple values of mean, median, and even standard deviation can be poor descriptors for the large data set. For this reason, many graphical distributions are presented in order to give context to what a large or small standard deviation or differing means and medians indicates.

One-way ANOVAs were primarily used to evaluate the a common independent variable on a dependent variable between two different categories of students (male/female, international/domestic, PHY 191/PHY 192, etc.) in order to determine if the category made a difference, however two-way ANOVAs were also looked at in order to try to determine the interactions within these categories. All p-values for these methods have been evaluated at the p = .05 significance level. Given the nonlinear nature of some of our data and the fact that there are hard lower and upper limits for grades and/or time data, thresholds had to be defined in some cases, in lieu of appropriate statistical measures, in order to determine what was high or low, by our arbitrary definition.

The other challenging feature to define was a trendline for results like the impact of the active time on the students’ final grades. A clear trend of diminishing returns was noted, and a general logarithmic trend line was fit, but only marginally better than a linear one. One major feature that exists is a trendline for the lower edge of the data, as can be seen in the relevant figures. Fitting a line to this first involves finding a suitable constraint to place on our logarithmic function:

\[ y(x) = a \ln(x) + b \quad \text{Eq. 1} \]

The initial method used to constrain this equation, in order to solve for \(a\) and \(b\), was to start by selecting a \(b\)-value for when \(x = 1\), that is the first term is equal to zero in Eq. 1. This choice could be anything, but some metric could be used to define this choice as well, like the lower quantile or highest grade of the bottom 5% of students of students as example. A second constrain could then be chosen for another \(x\)-value, in similar fashion to the first, and a value for \(a\) can then be found. For our purposes, we examined the grades of students who had mean active
times between 0.4 hours and 0.6 hours and between 0.9 hours to 1.1 hours, that is within ±6 minutes of 30 minutes and 1 hour and based our constraint on some percentage of their grade distribution.

While this process may seem arbitrary, this can be used to define minimum expectations for students, such as the minimum impact of reading at least 1 hour per assignment, or for finding students who follow a specific trendline that has been chosen, for example to find students that are potentially underperforming among their peers. In the case of finding the students that fit a specific trendline a method is suggested here, and implemented, as proof of concept. A line has been selected to fit the lower bound trend of student behaviors regarding mean active time’s impact on final grades by constraining \(a\)- and \(b\)-values as described above. The distance from each data point to a point on this line can be found by using the distance equation:

\[
D(x) = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad \text{Eq. 2}
\]

Where \(x_0\) is the mean active time for that data point and \(y_0\) is the final grade for that data point. For the \(x\)-values, a linear space was created from zero to the highest value of the mean active times used in steps of 0.001 hours. These \(x\)-values were then used with Eq. 1 to calculate the corresponding \(y\)-values and then the distance from each data point to each point along our line. Then for each data point the minimum distance created this way was stored. The \(R^2\) value of our data fit to this line is calculated and used as the constraining variable for which data points we will keep. If the desired \(R^2\) has not yet been obtained, you remove the data point furthest from the line and repeat the calculation until you have the desired range of students selected.

Lastly, the class page view times and number of page views data was addressed. The number of times a page was viewed per minute was an additional metric, besides the two separately, that was looked at. They were all evaluated for means and standard deviations. MATLAB was then used to find all the pages that were outside of defined bounds for time, number of page views, or number of views per minute. These results were used to inform a brief qualitative evaluation of the content on these pages. Each page outside of these bounds were examined in order to determine which broad category of content was on this page.

Viewing Time, Active Time and Number of Annotations

After initial statistical analysis and EDA it was clear that there was a fairly strong, positive correlation between viewing time, active time, and the number of annotations made, as will be shown. Active time indicated times that we knew students were interacting with their book, which included making their annotations, meanwhile viewing time included brief periods of inactivity or sessions where they allowed their browser to long out from 15 minutes of inactivity. In general, more viewing time results in more active time by a student and more active time yields a greater number of annotations. Given this fact and the very large variation in
Viewing Times between students, we have chosen to primarily focus on the impact of Active Time and number of annotations on the students’ grades for the course. This was done to avoid including these inactive periods. However, viewing time was used in comparison with active time to attempt to create a complete picture of how much time a student spent reading versus possibly typing.

Institutional Review Board

As this study involved the evaluation of human subjects in a classroom and online learning environment, an Institutional Review Board (IRB) application was submitted. The primary investigator for this application was Dr. Jennifer Blue. However, this study was classified as exempt from IRB review per 45 CFR 46 categories 1 and 2. That is to say that this study is considered to be part of the normal instructional practices and procedures for our institution and does not place the student at risk for criminal or civil liability. A copy of this exemption has been provided in Appendix B. That being said, once the data was collected all identifying information was removed to ensure the safety of all students involved in this study. Students we also given the opportunity to opt out of surveys, should they not want to be included, and non-identifying options for demographics were provided. Additionally, all participating students were 18 years of age or older.
DATA

Across all sections there was 2091 students worth of data available, however not all of these were unique students as some, but not all, were enrolled in both PHY 191 and PHY 192. Since the content and individual assignments varied and students did not always have the same section or instructor semester to semester, we have chosen to not compile data by student, but by section and demographic breakdown only. The PHY 191 sections included 1456 students and the PHY 192 sections included 635 students. One section of each semester included a set of honors students, 87 in total, who were integrated within the regular course. Most students took the traditional course layout with PHY 191 in the fall semester and PHY 192 in the spring semester, however 547 students took the trailer section of these classes.

Survey Responses

The survey that was given to students was voluntary and given the next class period after their second exam, which tended to have a particularly low turnout. As a result, demographic data was collected for less than 40% of students. Approximately 34% of students who responded identified as female (n = 330) and 66% identified as male (n = 640), with only two students out of the whole population identifying as other. Due to this very low number of students responding as other we have chosen to only look at the male and female portions of the data. Students largely identified as domestic (n = 768) when asked about their international status, but a significant portion did identify as international (n = 217). The students were also asked to identify their year in school, being either freshmen (n = 528), sophomores (n = 312), juniors (n = 117), or seniors (n = 31). Lastly, students were asked to identify their field of study with the options being STEM (n = 878), biology/microbiology (n = 29), kinesiology and health (n = 26), and outside of STEM (n = 44).

Examining the non-demographic survey data revealed that 91.4% of students had not used anything like Perusall but found it easy to start using. Approximately 50% of students felt their amount of time using Perusall met the professor’s expectations and/or their effort put into using Perusall matched their understanding of the material. Many students, 68%, felt using Perusall meant that they read the textbook, but still needed to interact with other students during class to aid in learning. When asked if they felt reading the textbook would help them understand their science classes, 81% felt positively it would. In terms of Perusall usage, 68% felt using Perusall helped them better understand their professor, while 62% felt it helped them with learning physics. This survey indicated that, overall, students felt using Perusall was a positive and beneficial experience.
Initial Statistical Evaluation

What follows are boxplots for active time, number of annotations, and final course grade, here after referred to as final grade, for all students involved, broken down by semesters and type of class, as well as divided by demographics. These have been included with the mean and standard deviations for each. The red X’s denote statistical outliers. These are presented here, void of commentary which has been placed at the end, for ease of comparison. Later graphs may or may not contain two sets of data or placed adjacent to one another, depending on the needs of the data being presented. While the graphs depict the relationship between variables, they are also presented in their entirety as an opportunity to looks at the distribution of data for each axis as well as in the broader sense.

**Boxplots and Means for All Students (n = 2091)**

![Boxplots and Means for All Students](image)

*Figure 1: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in this study.*
Boxplots and Means for Academic Year ’17-’18 (n = 1129)

Figure 2: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in academic year ’17 - ’18.

Boxplots and Means for Academic Year ’18-’19 (n = 962)

Figure 3: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in academic year ’18 - ’19.
Boxplots and Means for PHY 191 (n = 1456)

Figure 4: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in PHY 191.

Boxplots and Means for PHY 192 (n = 635)

Figure 5: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students in PHY 192.
Boxplots and Means for Female Students (n = 330)

Figure 6: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as female.

Boxplots and Means for Male Students (n = 640)

Figure 7: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as male.
Boxplots and Means for International Students (n = 217)

Figure 8: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as international.

Boxplots and Means for Domestic Students (n = 768)

Figure 9: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for all students who identify as domestic.
Boxplots and Means for Freshman Students (n = 528)

Figure 10: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for freshman students.

Boxplots and Means for Sophomore Students (n = 312)

Figure 11: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for sophomore students.
Boxplots and Means for Junior Students (n = 117)

Boxplots and Means for Senior Students (n = 31)

Figure 12: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for junior students.

Figure 13: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for senior students.
Boxplots and Means for Students in STEM (n = 878)

Figure 14: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in STEM.

Boxplots and Means for Students Outside of STEM (n = 44)

Figure 15: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students outside of STEM.
Boxplots and Means for Students in Health or Kinesiology (n = 26)

![Boxplots and Means for Students in Health or Kinesiology](image)

Figure 16: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in health or kinesiology.

Boxplots and Means for Students in Biology or Microbiology (n = 29)

![Boxplots and Means for Students in Biology or Microbiology](image)

Figure 17: The boxplots, means and standard deviations for mean active time, mean number of annotations, and final grade for students in biology or microbiology.
Overall, 92.6% of all assignments were attempted by students, meaning that it contained at least one annotation that scored 0 or higher. From Figure 1 we can see that, in general, students spend approximately 30 mins on the average assignment, making 4.5 comments in that amount of time. For context, the mean viewing time for all students was $0.87 \pm 0.85$ hours. This means that, on average, students spent 57% of their time active while on Perusall, though there is a large amount of variation in both of these times. For the entire course, for the average student, the total viewing time was $11.68 \pm 22.76$ hours, the total active time was $6.27 \pm 12.03$ hours, and the total number of annotations made was $55.4 \pm 92.8$ annotations. Again, these have notably high standard deviations. It is also worth noting that the median final grade for all students was 72.1, as median is a more robust than mean, but less informative with large populations where outliers are more common. The median mean active time for all students was 0.3633 hours.

When comparing academic years, students performed statistically similar in mean active time ($p = 0.119$), $0.50 \pm 0.47$ hours for the first year and $0.51 \pm 0.41$ hours for the second, but performed differently in mean number of annotations ($p = 0.0054$), $4.5 \pm 3.28$ annotations for the first year and $4.57 \pm 2.43$ annotations for the second, as can be seen in Figures 2 and 3. Keep in mind that the first year they were asked to make ten annotations per assignment and five were graded, but in the second year they were asked to make a number of annotations equal to the number of pages in the assignment, noting that most assignments were fewer than ten pages. The students also performed better in their final grades in the second year than the first year, an improvement of 8% ($p = 0.000002$).
In comparing PHY 191 with PHY 192, there is a statistical difference in the mean active time \((p = 0.0015)\), but no statistical difference in the mean number of annotations made \((p = 0.053)\) or their final grade in the course \((p = 0.081)\). Using Figures 4 and 5, we can see that students in PHY 192 completed their assignments in only 69% of the time it took PHY 191 students, and with less variance in these results, while maintaining the same number of annotations and final grade distribution. It is worth noting that PHY 191 is a prerequisite for PHY 192 and most of these students had had experience with Perusall by this point.

The trailer section students have been presented separately in Figure 18 as their Perusall assignments were weighted differently, 10% of their final grade instead of 5%. While ANOVA confirms that students in these classes performed differently in terms of mean active time and mean number of annotations \((p = 0.002\) and \(p = 0.007\) respectively), the final grades were not statistically different \((p = 0.17)\). A two-way ANOVA confirms that there are statistical differences between the various instructors and sections, however the trailer sections and instructor were not statistically different \((p = .121)\).

While male students and female student performed similarly in terms of final grades \((p = 0.15)\), this is the only area they performed similarly in. From Figures 6 and 7, we see that female students were much more likely to spend more time active in their textbook, \(0.74 \pm 0.53\) hours versus \(0.54 \pm 0.41\) hours \((p = 0.00003)\), and significantly more likely to make more annotations, \(6.2 \pm 3.55\) versus \(4.8 \pm 2.54\) \((p = 0.00000003)\). International and domestic students, however, appear to perform nearly the same on their Perusall assignments in all three areas, mean active time \((p = 0.086)\), mean number of annotations \((p = 0.051)\), and final grades \((p = 0.052)\), seen in Figures 10 and 11. As a reminder, all students were randomly assigned their groups in class as well as their groups in Perusall and so had ample opportunity to interact with individuals from all demographic groups.

Upon initial inspection of Figures 10, 11, 12 and 13, there appears to be significant difference between the freshman, sophomores, juniors, and seniors in each category, mean active time \((p = .0015)\), mean number of annotations \((p = 0.022)\), and final grade \((p = 0.005)\). However, when placed in two broad categories, underclassmen (that is freshman and sophomores) and upperclassmen (that is juniors and seniors), different trends appear. Underclassmen perform very similarly to one another in terms of mean active time, mean number of annotations, and final grade \((p = 0.63, 0.44, \text{and } p = 0.15, \text{respectively})\). Similarly, upperclassmen perform very similarly to one another in terms of these as well \((p = 0.64, p = 0.20, \text{and } p = 0.83)\). Every combination of underclassmen and upperclassmen, aside from these, results in \(p\)-values of less than 0.05, indicating significant differences in the performance of these groups. Lastly, from Figures 14, 15, 16, and 17, we find there was no significant difference between any of the groups by field of study for any of the categories \((p = 0.43, p = 0.17, \text{and } p = 0.10)\). Due to a lack of foresight when asking the students about their major, we realized that these categories did not represent the groups we were hoping to analyze. As such, no further examination of them will be presented.
Mean Viewing Time and Mean Active Time

![Graph showing the relationship between mean viewing time and mean active time.](image)

**Figure 19: The impact of mean viewing time on mean active time for all students.**

Analyzing the ratio of mean active time to mean viewing time yields results between zero and one. A zero represents a student who logged in, but next completed any activity and allowed their session to log out. A one represents a student who remained actively moving their mouse, scrolling, or typing the entire time they were logged in. This allows us to have a sense of how much reading is possibly taking place by the students and provides us another way to compare groups of students. When the number of annotations is taken into consideration, on a case by case basis, this can inform us of the habits of each student when doing their Perusall assignments, for example, gaming the grading system or not using their time or resources efficiently. In Figure 19 we can see that the mean active time to mean viewing time ratio for all students was 0.46 with a relatively strong, positive correlation, $R^2 = 0.77$. This means that viewing time and active time are a good metric for one another. It also indicates that students were rather consistent in how they performed in their assignments. A clear line can be seen along $y=x$ where students have hit the natural upper limit of this ratio. However, despite the ratio having a lower bound of zero, we see very few students in a range the range below roughly the $y = x/4$ line.
Figure 20: The impact of mean viewing time on mean active time for students in the trailer section.

The trailer maintained a similar active time to viewing time ratio as the general population, 0.43, which can be seen in Figure 20. However, this section performed even more constantly than the general population, with $R^2 = 0.83$. 
Figure 21: The impact of mean viewing time on mean active time for academic years ’17-’18 and ’18-’19.

Figure 22: The impact of mean viewing time on mean active time for PHY 191 and PHY 192.

When grouped together by academic year, students appear to behave much more similarly within their peer group, with higher R$^2$ values than the population, from Figure 23. It appears that the two years, however, have performed very differently from one another in terms of the active time to viewing time ratio. The first year, academic year ’17-’18, had a significantly higher ratio of 0.6 when compared to the population ratio or academic year ’18-’19, which had the lowest ratio of all demographics at .39. In Figure 22, this gap in ratio does not appear to be as prevalent between PHY 191 and PHY 192, but the PHY 191 class does appear to perform less consistently than they do the subsequent semester.
Figure 23: The impact of mean viewing time on mean active time for students who identify as male or female.

While male and female students appear to perform similarly in terms of their active time to viewing time ratio, females tend to be slightly more active than their male counterparts, seen in Figure 33. Likewise, international students also appear to be more active than their domestic counterparts, as well as more consistent in their habits, seen in Figure 24.
Figure 25: The impact of mean viewing time on mean active time for freshman and sophomores.

We see in Figure 25 that freshman students seem to be one of the most inconsistent of all the demographics, alongside domestic students in general. Meanwhile, in Figure 26, seniors appear to be the most consistent demographic and junior students appear to perform very similarly to freshman students.

Figure 26: The impact of mean viewing time on mean active time for juniors and seniors.
Mean Active Time and Mean Number of Annotations

Figure 27: The impact of mean active time on mean number of annotations time for all students.

Let us now take a brief look at the impact of mean active time on the mean number of annotations. It intuitively makes sense that the more active the student, the more likely they are to make annotations. This is complicated by the fact that assignment instructions set a predefined target for students in terms of how many annotations to make. As a result, the information is not as straightforward to evaluate. In Figure 27 it can be seen that the number of annotations made is impacted by the mean active time, but it also tends to cluster near the five annotations regions, with the exception of a sizeable cluster of students near the origin. Figures 28 and 29 have been provided as examples of how these instructions influence this data, as the two academic years had different sets of instructions. These are provided separately to give a clear, uncluttered picture of what is happening in each. The number of annotations in Figure 28 are much more tightly grouped for the first academic year, but both academic years see the majority of their comments coming before 1 hour of active time as elapsed.
Figure 28: The impact of mean active time on mean number of annotations time for academic year ’17-’18.

Figure 29: The impact of mean active time on mean number of annotations time for academic year ’18-’19.
Figure 30: The impact of mean active time on mean number of annotations time for PHY 191 and PHY 192.

While their trendlines suggest that they performed similarly, it is relatively clear to see by visual inspection of Figure 30 that the PHY 192 student was more efficient with their time, making a very consistent number of annotations in less overall time than the PHY 191 students. However, there also appear to be more possible instances of gaming the system, as cautioned about previously, in the PHY 192 population as well, with a good representation of their data points near the five annotations point, but with very little active time, keeping in mind that for PHY 192 n = 635, while for PHY 191 n = 1456. PHY 191 students appear to be far more likely to either take extreme amounts of time, make an extreme number of annotations, or both.
Figure 31: The impact of mean active time on mean number of annotations time for the trailer sections.

For the trailer section of physics students, two main groups of clustering can be seen on Figure 31. One is the larger, less tightly bound group of students who appear to gravitate towards producing an average number of annotations in an average amount of time. The other group tends to cluster very near the origin, suggesting a large amount of inactivity. Additionally, there are also appears to be a large number of students who produced numerous annotations in a small amount of active time, suggesting gaming the Perusall grading system.
Figure 32: The impact of mean active time on mean number of annotations time for PHY 191 and PHY 192.

The comparing the male and female trendlines from Figure 32 would suggest that, overall, female students perform similarly to male students when it comes to the rate at which they make annotations over time, but they consistently make more annotations in general. This is consistent with the information in Figures 8 and 9, that female students spend more time active in Perusall, compared to male students. Female students are much more likely to have large, extreme values in both mean active time and number of annotations as well. Lastly, it would appear that far fewer female students underperform or attempt to game the Perusall grading system.
Figure 33: The impact of mean active time on the mean number of annotations by international status.

Despite the apparent differences in trendlines, these two graphs appear to share some similar features. Both are scattered across the figure, with either no grouping, grouping around five annotations, or with a significant portion of their population near the origin. The domestic students do appear to have more extreme values in terms of number of annotations made than the international students and the international students have a similar trend in mean active time. This could be the reason for the difference in their trendlines.
Figure 34: The impact of mean active time on the mean number of annotations time by international status.

With so few upperclassmen represented (n = 148) it is difficult to evaluate any overall behavior by their demographic through visual inspection of Figure 34 when compared with the underclassmen, though a good number of individuals appear to have results near the origin. Beyond this, they appear to perform similarly to underclassmen, though with far fewer outliers. This could be related to the way they performed in terms of mean active time and mean viewing time.
Figure 35: The impact of mean active time on final grade for all students.

One of the biggest questions in this study is how does using Perusall impact the student’s final grades. A linear model does not appropriately approximate the results a student sees from being active in Perusall. It is visually apparent from Figure 35 that being active in Perusall does positively correlate to a better grade, albeit with diminishing returns. Diminishing returns is a well-known characteristic of a few functions and leads us to fit a logarithmic curve to the data.
Figure 36: The impact of mean active time on final grade for all students.

A logarithmic fit does help explain some of the variation in the data, as evident by the improved $R^2$ value in Figure 36 but does not fit the data as well as we’d like. It is apparent that the spread of grades that students earned is significant and that not all students equally benefited from increased active time. While a student can have a very small mean active time and still perform well in the class, there seems to exist a lower bound for students which suggests that greater active time equates to a greater minimum final grade among students. What follows is an attempt to quantify this lower bound, as described previously.
Figure 37: The impact of mean active time on final grade for all students with lower bound.

Figure 37 depicts a lower bound line which was defined by constraining the predefined regions, that is 0.5 hours and 1 hour, to two sample students. For the 0.5-hour constraint, a student was selected who earned a grade in the D-range for their course. The other student was chosen for having a final grade on the C+/B-cusp for their course. This means that every student who had at least this active time and who are above this trendline passed the course. For this reason, a method of selecting students who potentially are in danger of not passing has been developed, as a proof on concept for the use of this trendline.
Figure 38: The impact of mean active time on final grade for students fit to the lower trendline $y = 17.2671 \ln(x) + 68.55$.

Figure 38 shows only the students whose variation from our chosen trendline, that is $y = 17.2671 \ln(x) + 68.55$, results in an $R^2$ value of about 0.75. These students represent a potential group of students who are either in danger of not passing the course without intervention or who could potentially move into the category of above average students, in terms of final grades. This method could also be viewed as a way to remove outliers, both significantly above and below average students, for whom additional reading has limited impact. Now that we have identified these trends for the mean active time’s impact on final grade, the following figures will be shown with no trendline. The purpose of this is to, instead, focus on visually determining clustering behavior and the differences between the related demographics of students, which a poor fitting line will not add in.
Figure 39: The impact of mean active time on final grade by academic year.

It appears from Figure 39 that both academic years were well distributed in the use of their active time within Perusall and their final grades. It does appear, from visual inspection, that there were fewer times close to zero in the second year than there was in the first. There were also fewer extreme values of active time for the second year. It should also be noted that the second-year students did not have as many instances of low scoring individuals from approximately 1 hour of mean active time and onward.
Overall, it appears that PHY 192 students performed more consistently, forming a tighter cluster of students above the previously mentioned lower threshold, seen in Figure 40. There also appears to be far fewer extreme values and outliers for the PHY 192 students when compared with the PHY 191 students. Keep in mind that the majority of PHY 192 students had use Perusall the previous semester and all of them had passed PHY 191, or an equivalent course, in order to have taken PHY 192, so this may be reflected in this data.
Figure 41: The impact of mean active time on final grade for the trailer sections.

The trailer section of physics classes differs slightly from the general population in that their smaller mean active time, as noted in Figure 20, creates a more compact grouping of data points, seen in Figure 41. In fact, very few students spent up to an hour active in their textbook. This makes it challenging to determine what the effect of additional reading beyond the population mean has.
Figure 42: The impact of mean active time on final grade by gender.

While it appears from Figure 42 that these male and female students are both well distributed across their mean active times, note that there are far fewer instances of low mean active times for female students. As is supported by Figures 6 and 7, the mean active times for male students tend to cluster around 0.5 hours, while the female students appear to be distributed across a range of times from approximately 0.5 hours to 1.5 hours. It is also worth noting that the male students tend to have more extreme values in terms of grade distribution.
Figure 43: The impact of mean active time on final grade by international status.

In general, both international and domestic students appear in Figure 43 to perform similarly. Both are well distributed across the graph and both seem represented in extreme values. This is supported by the previous ANOVAs that suggested there was no statistical difference between the two demographics in how they performed in terms of Perusall use.
Figure 44: The impact of mean active time on final grade by class.

While the number of students for each demographic is significantly different, it does appear in Figure 44 that underclassmen are much more varied than upperclassmen. When upperclassmen participate consistently on assignments, active times of 0.25 hours to 1 hour, they tend to perform consistently in the course as well. They also tend to contain less extremes in terms of time values, but also less high final grades, compared to the underclassmen.
One additional way we can view the student’s performance on their assignments is by evaluating the mean number of annotations made per student and the impact this had on their final grade. It is apparent from Figure 45 that a linear fit does not work well for this data, but the reason for this is also rather apparent. A very clear cluster of the majority of students can be seen around an average of five annotations per assignment, with a mean of 4.1 ± 3.1. This trend existed largely for students who scored 50 or higher as their final grade. As a reminder, students in the first academic year were asked to make ten annotations per assignment, with five being graded, but this changed for the second year.
Figure 46: The impact of mean number of annotations on final grade by academic year.

Despite having two different sets of instructions, both academic years’ data appears to cluster around five annotations per assignment, seen in Figure 46. In fact, the second academic year, who did not have the five-comment minimum, appears to have less spread in its data in general, when compared with the first year’s data. Academic year ’17-’18 had a mean number of annotations equal to 3.9 ± 3.4, while Academic year ’18-’19 had a mean number of annotations of 4.3 ± 2.6, supporting this idea. There also appears to be significantly more first academic year students who underperformed in terms of number of annotations, and overall grade, than from the second semester.
Figure 47: The impact of mean number of annotations on final grade for PHY 191 and PHY 192.

It is apparent from Figure 47 that PHY 192 students, having used Perusall the semester prior, are much more consistent in their comment making and stick very tightly to the five-comment trend previously seen, with a mean of $4.2 \pm 2.5$ annotations per assignment. As has been the trend with PHY 191 students, they remain more spread out in terms of number of annotations and grade, having made $4.7 \pm 3.0$ annotations per assignment on average. There appears to be, in general, far fewer outliers for the PHY 192 students.
Figure 48: The impact of mean number of annotations on final grade for trailer sections.

Though there are data points distributed fairly broadly in Figure 48, the students of the trailer sections tend to group around the four and five annotations region, even though the average for the classes were $3.8 \pm 1.9$ annotations per assignment. Additionally, students who reached this number of annotations tended to pass the class. In this way, they seem to perform in similar fashion to the general population.
Figure 49: The impact of mean number of annotations on final grade by gender.

While in Figure 49 female students still tend to cluster, they tend to do so higher in number of annotations than male students, 6.0 ± 3.7 versus 4.7 ± 2.7. Female students were also more likely to make a much greater number of annotations, but less likely to make far fewer annotations.
Figure 50: The impact of mean number of annotations on final grade by international status.

Domestic students have a mean number of annotations per assignment of $5.3 \pm 3.1$, from Figure 9. International students, by comparison, made $4.6 \pm 2.9$ annotations per assignment, from Figure 8. The two groups did perform similarly, as their ANOVA suggested, however international students did have more students per capita who fell below a final grade of 50 when making near five annotations on average, which can be seen in Figure 50.
Though there are far fewer upperclassman, they tended to perform more like the general student population in terms of mean number of annotations, with $4.5 \pm 2.6$ annotations per assignment. From Figure 1, we note that the whole student population had a mean of $4.5 \pm 2.9$. In contrast, even though they consisted of the majority of the population, underclassman had a far more varied amount of annotations made, $5.3 \pm 3.2$ per assignment, which also tends to be true of their final grades. However, you can see in Figure 51 that they still maintained the trend of clustering around five annotations per assignment.
Modification of Perusall Settings

One advantage to Perusall’s settings is that they can be applied retroactively. This allowed us to take an already completed class and modify certain settings to see how it would have impacted the students, were we to use this setting in the future. Two major changes were implemented, separately, to evaluate their effect: distribution penalty and annotation score. This was done on a class-by-class basis and each evaluated separately, in the event that any one class was impacted far more greatly than the others. However, this was not the case overall as all ANOVAs within a class showed no statistically significant difference (p > 0.05). New means and standard deviations were then calculated and the number of annotations these changes impacted were counted.

For the distribution penalty, the penalty was turned to 100% and the grades recalculated. This change impacted about 20% of all assignments, however it only resulted in an 8% drop in mean assignment grades, from 4.48 per assignment to 4.1 per assignment, though there was an almost 30% drop in the median grade for Perusall assignments with this change, from 5.5 per assignment to 4 per assignment. With this result, smaller changes in distribution penalty were not attempted, as the impact would have been even less significant. Standard deviations were statistically unaffected by this change.

For the grading system, the 0/1/2 scale for annotation scores was replaced with a 1/2/3 scale in order to provide students with a participation score for just making any comment, regardless of quality or length. This resulted in no statistically significant difference in mean grades, less than 1% (p = 0.337), and maintained the same median grade. Standard deviations were statistically unaffected by this change.
Page View Statistics

![The Number of Page Views by Page Number](image)

Figure 52: The number of times each page was viewed by each class by page number.

Overall, each page that was assigned was viewed an average of 137.4 ± 105.5 times over the course of the class. However, as the standard deviation of this value implies, there was a significant amount of variation in the number of times each page was viewed. As would be expected with any technology that includes scrolling, scroll depth is a factor in how many times a page is viewed. Pages at the beginning of an assignment are significantly more likely to be viewed multiple times, while pages at the end of the assignment are much less likely to be reached when each student visits each time. Similarly, pages at the end of an assignment are much less likely to be viewed. As a comparison, the average number of times a page was viewed if it was the first or second page of the assignment was 233.9 ± 116.6, but the average number of times a page was viewed if it was the last two pages of an assignment was 121.3 ± 113.8, again the large standard deviation indicating there is a lot of variation in this. In general, the number of page views tends to be higher at the beginning of the semester, noting that PHY 192 begin their assignments around page 490, depending on instructor, a feature apparent in Figure 52. Additionally, the extra credit assignment given in PHY 191 begins on page 1063.
Figure 53: The average amount of time each page was viewed by each class by page number. This data has excluded three extreme values for ease of viewing: 278 minutes for page 26, 168.69 minutes for page 624, and 146.06 minutes for page 850.

Each page was viewed an average of 3.16 ± 4.28 minutes across all assignments. From Figure 53 we can see a slightly similar trend at the beginning of an assignment, with 2.86 ± 1.39 minutes being the average for the first two pages of an assignment and 2.23 ± 3.29 minutes for the last two pages. Of note here it that the variation in times is significantly different for the last two pages than the first two.
The average amount of time each page was viewed by how many times the page was viewed by each class. This data has excluded three extreme values for ease of viewing: 278 minutes for page 26, 168.69 minutes for page 624, and 146.06 minutes for page 850.

The average length of time spent on each page is not correlated with the number of times the page is viewed, as seen in Figure 54. That being said, it appears as though the more time a page is viewed, the less likely it is that that page will have an exceptionally high average viewing time. However, it also appears that the more times a page is viewed, the less likely it is to have a near zero average viewing time as well.

The use of each page by each individual section was evaluated as well. It should be noted that ANOVAs were performed in regard to both number of pages views and mean page times between all of the sections. A statistically significant difference was present for every metric between the sections with the except of the trailer sections, whose number of page views and mean time were statistically consistent (p = 0.214 and p = 0.067), and four sections of PHY 192, whose mean time was consistent (p = 0.482). These ANOVAs did not take into consideration differences in assignment length or individual pages assigned.

The standard deviation of each metric was then used to select the most and least viewed pages for study. This presented an issue, since both time and number of page views could not be negative. Instead, an artificial lower bound was established for each, as needed. When looking at the most viewed pages of the textbook, there was an overwhelming trend of the early
assignments for the semester having an exceedingly high number of page views. Within the first 200 pages, 48 pages had page view numbers that exceeded the mean by at least one standard deviation, that is greater than 242.9 views, in at least 75% of PHY 191 classes. For PHY 191, the first eleven pages of an assignment were most likely to contain a page that was viewed more than one standard deviation above the mean, occurring 95.2% of the time. Similarly, PHY 192 was most likely to occur within the first four pages. This was also the most likely time for pages to be viewed an extreme number of times, with an average of 6.6 pages per section with page views more than three times the standard deviation above the average, that is more than 453.9 views, all falling within these first few pages. Average times maintained similar patterns, but at much reduced rates and sizes. In general, times that exceeded the mean by one standard deviation we distributed throughout the book but did tend to happen most on the first full page of an assignment, often neglecting the chapter title page, occurring in at least 50% of the sections.

The first standard deviation for number of page views could be used to establish a lower bound, which was 31.8 views, but additional applications of the standard deviation was not visible. Similarly, no lower bound could be found this way for page view times. As such, further lower limits were defined to be 10 page views or less than 6 seconds of viewing time. The least viewed pages in terms of time and number of page views varied much more widely than the most viewed, but they all came further into the assignments. Cross referencing numbers from each list of page views and viewing times that were less than our defined bounds, we found the 117 pages that had been assigned were significantly under viewed. These pages were examined to determine what kind of content each contained: 108 pages consisted of primarily worked example problems or the chapter summary, 6 pages contained a section called “Mathematical Aside”, 1 page was a table about the moment of inertia of objects with uniform density, and 1 page was the beginning of the simple harmonic motion section.
DISCUSSION

General Results

The behavior of students in regard to textbook and Perusall use varies widely, but there are a few broad statements that can be made. While there are students who do not benefit as much from interacting with their textbook or who need less time with their textbook to perform well, there are some metrics which seem to indicate that a student will perform well. As can be seen by the lower bound fitting in Figure 37, students who spend at least 30 minutes active, and possibly spend as much as twice as long reading than this based on the impact of viewing time on active time from Figure 19, were very likely to at least pass the class. Similarly, students who spent closer to an hour reading per assignment were likely to do above average with regards to the class average. It also appears, from Figure 45, that students who made approximately five annotations per assignment were much more likely to pass the class.

While the two academic years performed similarly in active time, there were several notable differences. The academic year ’18-’19 students were more consistent in the number of annotations they made and got closer to making five annotations per assignment than the academic year ’17-’18 students, which can be seen in Figures 28, 29, and 46. They also had a lower active time to viewing time ratio, which could imply more actual reading took place, as seen in Figure 21. There was no significant difference in their active time, and this appeared to have no impact on final grade. This is significant because of the major change in how the assignments were given, in terms of the number of annotations grades, ten versus half the page length of the assignment. It’s possible that it was just coincidence that half the number of pages led to a very similar output in terms of number of annotations, but it appears that varying the number of annotations required is a better pedagogical method.

The most significant quality of note between the PHY 191 students and the PHY 192 students is how much more efficient and concise the PHY 192 students appear to perform. This is very likely due to having used Perusall in PHY 191, and having passed that course, and is what we would hope and expect to see. This consistency and familiarity may be beneficial to instructors, who can make greater use of the confusion report generated by Perusall and target their lesson plans to these students. It also may prove to be a beneficial group to institute other methods of presenting Perusall assignments, such as those mentioned in the Implications for Instruction.

It is important to discuss the impact that percentage of final grade has on students reading and Perusall interaction as well. The trailer section’s final course grades were worth 10% of their final grade, compared to the 5% for the other sections. Notably, the students in this class tended to have more central values in regard to their active time to viewing time ratio, as seen in Figure 20. This could be considered ideal behavior, depending on how they spent that time annotating and reading, as fewer extreme values suggest a better, or at least mixed, use of their time. While
there does appear to possibly be students who are gaming the Perusall scoring system, as seen in Figures 30 and 31, it is inconclusive how the rate of these occurrences compared to the general population without a qualitative study as well. The results from both Figure 43 and Figure 50 do show promising results in terms of the impact of students’ behavior in Perusall with their final grade for the trailer section. The students appeared to maintain a high rate of passing the course with even less time spent active than the population, meanwhile students appeared to maintain at least the same rate of annotations production, if not more.

Comparing viewing time with active time gives us a good indication of how the students have spent their time while in Perusall, in general. While we have no defined goal for what the ratio of active time to viewing time should be, we can say that there are three main categories of students based on these ratios. If a student has a very low Active Time compared to their viewing time, these students are at risk for not completing their assignments in a timely or reasonable manner. Evaluating students with a ratio of 1/14, this is doing the bare minimum to avoid being logged out thus any student with a ratio greater than this would be subject to long periods of inactivity. Students with a ratio approaching 1 is also a possible problem. While this much active time could be a result of idly scrolling while reading, a student with this ratio, combined with an overall low viewing time and/or minimal number of annotations could be finding alternative ways to gain their Perusall grades, including such things as making exceptionally long annotations or copying and pasting text into Perusall. Lastly there are the students who are closer to a ratio of 1/2, but with a large variation either way. These would seem to model the typical student in most cases and may be a starting point in finding the ideal time, though individual traits, such as reading speed and comprehension, will cause this to vary significantly.

Given this though on active time and viewing time ratio, a trend to consider when looking at this data based on how many annotations were suggested to be made by the student, how many annotations were actually graded, and considerations for students who may have found creative solutions to getting their annotation points, as was cautioned. It can be noted in Figure 27 that there is a collection of students who made numerous annotations in a relatively quick amount of time. While we are not accounting for what grade those annotations received in this graph, these could potentially be students attempting to game the system. Similarly, in Figure 19, the previously mentioned upper limit for the active time to viewing time ration of 1 is possibly an indicator of students who entered Perusall, entered text without reading any significant portion until they received the desired score, and then exiting. This is especially true of students with low overall times, but an average or high number of annotations. Sorting by these students could be a method to avoid gaming the system or potential academic dishonesty. Unfortunately, no formal qualitative analysis has been done on these types of annotations and students yet.

It does appear that students are willing to commit 15 to 30 minutes of active time regularly on assignments, spending about 3 minutes on each page of the assignment. We can also infer from the number of times students viewed pages in the textbook that they enter their assignment multiple times during the assignment. Also, as a reminder, the students were able to see their assignment score as they made comments and adjust accordingly. If we consider that students need to spend time writing annotations, in addition to reading, it could suggest that
students are not actually reading all of what they are being asked to read, but simply entering their assignment long enough to make a significant number of annotations, with sufficient scores, to get their credit and be done. While this does not ensure students are reading, this does ensure students are interacting with their text and the subject matter on some level. Further study would be required to determine actual levels of reading and comprehension, as well as more granular studies of engagement or a qualitative analysis of this engagement by students in these courses.

It should be noted that many of these trends may not be a result of reading, specifically, but an indicator of engagement in the course. Students we are more likely to engage with their Perusall assignments may also be more likely to be engaging with their homework, attending lectures, and participating in student driven sections. In this way, metrics like these and the lower bound data selection presented in Figure 38 could be used as an indicator as to which students would benefit from additional engagement. This information could be used to select the students who could benefit from additional reading, but could also benefit from seeking tutoring, attending supplemental instruction sessions, and being given additional resources or targeted material.

Demographic Results

Noticeable difference: Students who identified as male tended to perform slightly better than those who identified as female, on average, but had much great variance in their performances. Female students tended to perform much more consistently, with both their final grade and active times generally falling into a smaller region. Female students are much more likely to spend more time active in Perusall and make more annotations than their male counter parts. The number of annotations/active time ratio of males and females, seen in Figure 30, is very similar, though the intercept value differs, and female students have a higher active time/viewing time ratio, seen in Figure 23, so this greater number of annotations is likely a direct result of more time spent in the book. In regard to final grades, despite being similar in performance, the impact of mean active time and mean number of annotations on this final grade differs for both males and females as well. In both Figure 42 and Figure 49, female students to distribute their active time and annotations beyond the mean value more often, however it is not clear that this provides any additional impact on their grade. This could suggest that female students are over performing, in terms of Perusall assignments, but only gaining in consistency from it and not in overall score. Male students vary more in nearly every metric, so it is possible that this also suggests that more engagement from male students could result in more consistent, if not higher, grades as well.

Despite being a language-based project, international students do not tend to differ from domestic students, in terms of Perusall assignment performance in most metrics. The active time to viewing time ratios of the two groups do appear to vary despite there being no statistical difference, as can be seen in Figure 24, but if p-value is any indicator then these two groups could potentially be influenced differently and quickly become statistically different, noting that
if we evaluated at a higher p-value level of significance, they would not be considered the same. For example, in Figure 43 they appear to have the same impact from active time on their final grade, however, it does appear that international students tend to spend less active time, but there are more outliers present as well. Another pitfall to be mindful of is that while on average international students perform similarly, the habits of some individuals leave them in danger of performing poorly for the course. In Figure 48 it can be noted that there was a higher percentage of international students who performed poorly in the overall class but did manage to consistently produce about five annotations per assignment. It is possible that completing their Perusall assignments before class helps international students grapple with language-based issues and verbiage before attending lecture, but those who are simply completing their assignments to earn points may not be receiving the same benefit as those who are actively reading and earnestly participating. However, this cannot actually be known without further investigation.

Overall, upperclassmen tend to perform more consistently than underclassmen, though they also tend not to be found on the extremes of most metrics either. This could be the result of a number of factors. Upperclassmen have more experience at being a student, including knowing their study habits, managing their schedules and social life etc., and this may be reflected in this. Some of these upperclassmen may also be repeating this class, but without knowing who they are we cannot account for that. The upperclassmen also tend to differ by major, which is known within the department, as some majors require PHY 191 or PHY 192, but it is not a prerequisite course for them. This could also be true of upperclassmen who are pre-med or considering medical school. Conversely, many freshmen and sophomores, especially those who identified as STEM, may have these courses as prerequisites for many of their other courses, especially for majors like engineering and physics. This is possibly why there is such variation in the underclassmen data, by comparison, more so than in any other group. This is ripe for further study.

Page View Results

Everything considered, it is apparent from Figures 52 and the general statistical evaluations of the impact on page views based on placement that students are likely to view pages multiple times if they are near the beginning of an assignment or earlier in the semester. This is especially true of PHY 191 students when compared with PHY 192 students. This result for being placed at the beginning of an assignment makes sense as scroll depth is an issue that needs to be considered when looking at number of page views. Large viewing numbers at the beginning of the semesters could be for a number of reasons, including students getting comfortable with their assignment, their study habits and schedule, or simply learning to more effectively earn their points in Perusall. This can account for the difference in page view numbers from PHY 191 to PHY 192, but the trend of declining number of views as the semester goes on still exists. However, in general, students should still be encouraged to ensure they are viewing the whole assignment.
Page viewing times are far more erratic than number of page views, as can be seen in Figure 53. There is no apparent trend in the relationship between page number and average viewing time, nor is there one between number of page views and average viewing time. Based on the brief qualitative study conducted we believe there may be trends that exist for page view times, as well as number of page views, based on the content of those pages, but a more in-depth study will be required. However, when cross referencing the lowest the lowest values of each category, we find at least one standout category of pages that are underutilized by students in their assignments: pages that contain example problems or the chapter summary. Since there is, at most, 42 chapter summaries in the edition of the book used for this course and not all of those chapters were used, there is a considerable number of worked example problems that were not viewed by the students. If an instructor’s hope was that the students would learn problem solving or critical thinking skills from these sections, then it is unlikely that that is the case for these particular pages.

Implications for Instruction

One large takeaway from this study is that instructors should be mindful of how students may use their time in Perusall and take appropriate actions. There does appear to be several instances of students attempting to game the system, or possibly cheat, as can be seen throughout the sections Mean Viewing Time and Active Time Mean Active Time and Mean Number of Annotations. Having the instructors of Teaching Assistances actively checking the assignments once a class period and flagging annotations that have received their score inappropriately could greatly deter this behavior. Additionally, importance of ensuring the syllabi for each class contains very specific language about what this type of annotation looks like and the consequences for making these types of annotations cannot be understated. This may also improve with higher impact on the final grade, as those students who lose points for making these types of comments will find it impacting their grade more.

Given that the initial pages of an assignment will always be viewed the most and certain sections are being viewed far less in both page views and time, one option to consider is simply being more judicious in the selection of pages. If students will be doing their critical thinking outside of the text, for example during homework, then triaging these pages may encourage more interaction on the topics one wishes to discuss in class specifically. This is could be especially useful if you are concentrating on developing specific language or math techniques to build on, giving the students a very formal and focused introduction from which to learn the application come lecture or homework.

Given these extreme differences between the most and least viewed pages for each given assignment, instructors may want to consider using more targeted methods of presenting the same content. One possible approach could be to use pre- and post-reading assignments. This could allow an instructor to present vital elements that would benefit the student during lecture, such as new terminology or ideas, in a pre-reading assignment. Then they can provide the
students with a post-reading assignment after lecture that focuses on problem solving skills that will be needed to complete their homework or aid in their studying. This would contrast greatly with what was seen in this study, where these critical thinking areas, including worked examples, were the least viewed pages of the book.

While instructors were told not to participate in Perusall, future methods of implementing Perusall assignments could benefit from the instructor “seeding” questions and comments in the assignment before the students begin making their own annotations. This could be used to guide students toward problem areas, increase conversations about particularly challenging or thought-provoking subjects, or encourage interaction with more of the textbook. If this method is used it would be important that the instructor not continue to engage with the students and the text beyond this initial seeding of questions. This is to prevent the instructor from unintentionally ending conversations between students by being seen as the authority and subject matter expert.

In looking at possible alternative grading systems for Perusall, it may appear odd that grades had no variation after changing from a 0/1/2 grading scale for annotations to a 1/2/3 grading scale for annotations. This could be for a few reasons and does not suggest that this method couldn’t work. Bear in mind that this new grading scale was applied after students had completed their assignments and so they were not intending to gain just a participation grade. Nor were they aware that the weights of each grade would differ. In a 0/1/2 scale, a final Perusall grade of 0 would be with 0% of their grade, a 1 would be worth 50%, and a 2 would be worth 100%. But, in a 1/2/3 scale, 1 would be worth 33% of their Perusall grade, while a 2 would be worth 67%, and a 3 would be worth 100%. So, students would be able to still receive a “passing” score with less effort. It may be worth considering using a modified version of this scale, such as 1/2/4, instead. This would maintain the ratio for the two highest scoring categories, still forcing students to make at least some good quality comments in order to receive a “passing” score, while allowing students to get credit for participating at all, albeit at the much lower rate of 25% of their grade.

In similar fashion, the grade distribution penalty is a possible area instructors could consider making a change to influence the behaviors of their students. While the impact of the distribution penalty on the students’ grades can be quite dramatic, students tended to either make enough comments distributed well enough so as not to endure the penalty, regardless of severity, or simply not do the assignment. This experimental change resulted in a negligible impact on the students’ overall grades, only 8% of their Perusall grade which was only approximately 5% of their overall grade at lowest. This suggests that turning this penalty up will not adversely impact students who are already performing well and maybe possibly serve as a motivator for those who are underperforming slightly. However, since these changes were tested on the grades of a class which already passed, further investigation into its use is needed to determine what impact it might actually have if used. Instructors who feel they need to encourage students to participate across the text more, and do not wish or cannot encourage to do so in other ways, could make use of this feature, whether at 100% or a lesser value, without fear of drastically impacting the students’ ability to accomplish their work.
Future Work

This study leaves many avenues of approach available for future study. As this was an initial study of the use of Perusall in these classes, it sought to quantify the use of it by the students and our methods of deployment. From here, nearly every metric or subject discussed could be extended and built upon. From a quantitative standpoint, many factors could be looked at more closely or with more rigor and some metrics could benefit from long term study. This can include introducing different techniques and methods of pattern recognition to the various data sets or examining the possible intersections of many of these variables, both of which may be possible with techniques such as machine learning. Controlling for instructor or creating a control group which does not use Perusall with whom data such as final grade can be compared with would be beneficial as well.

Something that was not able to be measured but could have impacted this study was the students’ use of outside material, such as other books, websites, or instructor’s notes. It would be possible to include supplemental material, such as instructor’s notes if normally provided, in Perusall as well. Even if the students do not annotate these, general time and page view data can be collected for comparison with the textbook and to help gauge study habits. Combining this with information on the date and scores of exams could help inform the use of these materials in students’ preparation for the course material.

This study could benefit from a more rigorous and complete survey of the students, especially demographically. Very specific groups of interest should be asked to identify themselves, if possible. For example, by major it may be useful to identify the engineering majors, pre-med students, mathematics majors, computer science majors, or more specific groups outside of these that may not be as well represented. Each of these categories of students has different needs and considerations and may prove beneficial to study even beyond PER, but for the department and university reasons. It would also be beneficial if students were given the survey when more students are available to take it, considering out smaller sample size when compared to the number of students’ data available.

Pre and post testing of students could also form a way to evaluate the impact of students’ reading and Perusall use compared to known data. One possible way to incorporate this is using a standard test, such as giving the FCI, at the beginning and end of PHY 191 as an assessment of the students’ conceptual understanding. If evaluating for something more specific, such as information recognition on a page, a similar standardized test could be used or even developed to help determine if this form of social annotation alters the way students perceive printed information.

Additional data could be included as well, to create a clearer picture of student behavior and what influences their final performance in a class. This can include traditional things that are already measured, such as exams, homework, and attendance. Perusall also allows us to track the time the users were active and, if the resources are available and able to be put to this research, additional information about the students’ use of Perusall, including a more specific break down
of time information, is available. If other trackable information exists, such as how students perform in and with Learning Catalytics and Mastering Physics or a Learning Management System, such as Canvas, that can also be included. This would allow for a complete picture of students habits in class and not only allow for richer and more in-depth research to be conducted, but also allow us to produce a model of student behavior and target instruction or intercede as needed when they are in danger of performing poorly. This could be used to identify students who may benefit from visiting an instructor’s office hours, supplemental instruction, tutoring, additional materials or online resources, or even just extra reading.

Beyond these quantitative approaches, much can be done with qualitative analysis as well. Using some of the information provided here could aid in a qualitative examination of the actual annotations students have made. These can be looked at from numerous perspectives, including the type of annotation made such as questions, replies, and off-topic comments, the interaction between students, or the type of material annotated in the text. These qualitative approaches could be used to identify beneficial or problem areas or help generate metrics to prevent or encourage certain behaviors.

Lastly, many of these assessments could be combined with new information to evaluate the use of the textbook in general. For example, if combined with eye tracking devices, information could be collected on how various sections are utilized, especially in conjunction with the previously mentioned study of annotations and/or alongside metrics defined here, such as active time to viewing time ratios. This assessment would not be used for determining what a typical student does when using Perusall, but for evaluating how the student interacts with the content, in terms of the content itself, how that content is used to develop the language to converse with, and to help define what good metrics should be for evaluating students’ performances.
BIBLIOGRAPHY


APPENDIX A: PERUSALL SURVEY

Please record all answers on the testing form provided.
Please write (and bubble in) your unique ID on the testing form.

Perusall Project Survey

1. Have you used Perusall before you started taking this class?
   a) Yes
   b) Not sure
   c) No, but I’ve used something like it
   d) No

2. It was easy to start using Perusall.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

3. The time I spend using Perusall meets my professor’s expectations.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

4. My effort using Perusall reflects my understanding of the materials.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

5. Using Perusall helps me discuss physics more with classmates outside of class.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree
6. Using Perusall means I no longer need to discuss physics with classmates during class.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

8. Even in science classes without Perusall or anything like it, I always read the book.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

9. Reading the book in my science classes would not help me learn.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

10. Using Perusall helps me learn physics.
    a) Strongly Agree
    b) Agree
    c) Disagree
    d) Strongly Disagree

11. Using Perusall does not help me understand what the professor says during class.
    a) Strongly Agree
    b) Agree
    c) Disagree
    d) Strongly Disagree
12. I do not need to read the book since the professor goes over everything during class.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

13. I wish my other classes would start using Perusall.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

14. I usually review the instructor’s Power Points while studying or doing homework for this class.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

15. I do not read or watch anything other than the textbook or instructor Power Points while studying or doing homework for this class.
   a) Strongly Agree
   b) Agree
   c) Disagree
   d) Strongly Disagree

16. What is your year in school?
   a) First year of college
   b) Sophomore
   c) Junior
   d) Senior
   e) Other / Prefer not to answer
17. What is your major?
   a) Biology / microbiology
   b) Kinesiology and health
   c) Another science, engineering, or math (STEM) major
   d) A major outside of STEM
   e) Prefer not to answer

18. What is your gender identity?
   a) Female
   b) Male
   c) Other
   d) Prefer not to answer

19. Do you consider yourself an international student?
   a) Yes
   b) No
   c) Prefer not to answer

20. Are you at least 18 years of age?
   a) Yes
   b) No

21. Do we have your permission to use your answers in our research study?
   a) Yes
   b) No
To:      Jennifer Blue (bluejm@miamioh.edu)

Physics

Re:      Investigation of student attitudes and achievement in an electronics course

Project reference number is: 02652e
(please refer to this ID number in all correspondence to compliance administration)

The project noted above and as described in your application for registering Human Subjects (HS) research has been screened to determine if it is regulated research or meets the criteria of one of the categories of research that can be exempt from approval of an Institutional Review Board (per 45 CFR 46). The determination for your research is indicated below.

The research described in the application is regulated human subjects research, however, the description meets the criteria of at least one exempt category included in 45 CFR 46 and associated guidance.

The Applicable Exempt Category(ies) is/are: 1/2

Research may proceed upon receipt of this certification and compliance with any conditions described in the accompanying email message. When research is deemed exempt from IRB review, it is the responsibility of the researcher listed above to ensure that all future persons not listed on the filed application who i) will aid in collecting data or, ii) will have access to data with subject identifying information, meet the training requirements (CITI Online Training).

If you are considering any changes in this research that may alter the level of risk or wish to include a vulnerable population (e.g. subjects <18 years of age) that was not previously specified in the application, you must consult the Research Ethics & Integrity Program before implementing these changes.

Exemption certification is not transferrable; this certificate only applies to the researcher specified above.
All research exempted from IRB review is subject to post-certification monitoring and audit by the compliance office.

Best of luck with your research,
Jennifer

--

Jennifer Sutton, MPA, CIP
Associate Director of Research Ethics & Integrity
Office for the Advancement of Research and Scholarship
Miami University
102 Roudebush Hall
Oxford, OH 45056
Phone: 513-529-0454
http://www.MiamiOH.edu/compliance