January 2016

THE DEVELOPMENT OF A COMPUTERIZED INTERACTIVE TEACHING ASSISTANT IN PHYSICS: THE CITA ON CHIP PROJECT

Cyrus Marazban Vandrevala
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Cyrus Marazban Vandrevala

Entitled

THE DEVELOPMENT OF A COMPUTERIZED INTERACTIVE TEACHING ASSISTANT IN PHYSICS: THE CITA ON CHIP PROJECT

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Andrew S. Hirsch
Co-chair

Lynn A. Bryan
Co-chair

Laura J. Pyrak-Nolte

Hisao Nakanishi

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Andrew S. Hirsch and Lynn A. Bryan

Approved by: John P. Finley — 09/02/2016

Head of the Departmental Graduate Program — Date

THE DEVELOPMENT OF A COMPUTERIZED INTERACTIVE TEACHING ASSISTANT IN PHYSICS: THE CITA ON CHIP PROJECT

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Cyrus M. Vandrevala

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

To my family and friends for their support, and to the student who started me down this path.

ACKNOWLEDGMENTS

This research project was supported by a grant from the College of Science at Purdue University. I would also like to thank the PER@Purdue research group for their help over the past couple of years. The group has supported me in a number of ways including: sharing their analysis techniques, introducing me to members of the PER community, providing equipment for my research interviews, and offering good advice when I was unsure of how to continue. Thank you all.

PREFACE

All of the work presented henceforth was conducted at Purdue University in West Lafayette, IN. The studies and associated methods were approved by the Internal Review Board at Purdue University in October 2015 (IRB Protocol #1509016549). Multiple chapters in this thesis are supplemented by research papers that are in pre-print (soon to be published).

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	xii
ABBREVIATIONS	xvi
GLOSSARY	xvii
ABSTRACT	xviii
1 INTRODUCTION	1
1.1 Background	1
1.2 Overview of Electricity and Optics	2
1.3 The CHIP Homework System	3
1.3.1 Strengths of CHIP	5
1.3.2 Weaknesses of CHIP	6
1.4 The Need for CITA	7
1.5 Research Statement	9
1.6 Organization of the Thesis	11
2 THEORETICAL FRAMEWORK AND LITERATURE REVIEW	13
2.1 Background	13
2.2 Constructivism	14
2.3 Polya's Problem Solving Strategy	17
2.3.1 Understand the Problem	17
2.3.2 Devise a Plan	17
2.3.3 Execute the Plan	18
2.3.4 Review the Solution	18
2.4 Problem Solving in Physics	18
2.4.1 Novice Problem Solvers	18
2.4.2 Expert Problem Solvers	20
2.5 Scaffolding in Physics	21
2.6 Multimedia Learning	23
2.6.1 Cognitive Theory of Multimedia Learning	24
2.6.2 Mayer's Twelve Design Principles	24
2.7 Existing Online Homework Systems	30
2.7.1 WebAssign	31
2.7.2 McGraw Hill Digital Platforms	32
2.7.3 Sapling Learning	32

Page

2.7.4 Mastering Physics 33
2.7.5 The Expert TA 33
2.7.6 LON-CAPA 34
2.7.7 FlipIt Physics 35
2.7.8 Minds on Physics Learning Modules 35
2.7.9 Customizable Computer Coaches for Physics Online 36
2.7.10 Learning Management Systems 37
2.7.11 Massive Open Online Courses 38
2.8 Future Directions 38

3 METHODS 41
3.1 Preface 41
3.2 Research Design 41
3.3 Research Participants 42
3.3.1 Previous Skills 43
3.4 Design of the CITA System 44
3.4.1 Structure of the Tutorials 44
3.4.2 Schedule of Development 47
3.5 Quantitative Procedures 48
3.5.1 Statistical Methods 48
3.5.2 Exam Scores 50
3.5.3 Brief Electricity and Magnetism Assessment 51
3.5.4 Multi-Step Problem 51
3.5.5 Student Exit Survey 52
3.5.6 Recording Clicks 52
3.6 Qualitative Procedures 52
3.6.1 Strategy of Inquiry 52
3.6.2 Focus Group Sessions 53
3.6.3 Analysis Procedures 54
3.6.4 Roles of the Researchers 54
3.7 Quantitative and Qualitative Connections 55
3.8 Data Collection Procedures 56

4 OVERVIEW OF THE CITA SYSTEM 57
4.1 Preface 57
4.2 Introduction 58
4.3 Description of the Class 59
4.4 Literature Review 60
4.5 Design of the CITA System 63
4.5.1 Structure of the Tutorials 63
4.5.2 Schedule of Development 67
4.6 Methods of Analysis 69
4.6.1 Quantitative Procedures 69

Page
4.6.2 Quantitative Data Sources 69
4.6.3 Qualitative Procedures 71
4.6.4 Focus Group Sessions 74
4.7 Results 75
4.7.1 Usage of CITA 75
4.7.2 Student Improvement 80
4.7.3 Student Opinions 99
4.8 Discussion 105
4.9 Conclusions 107
4.10 Acknowledgments 109
4.11 Appendices from the Paper 109
4.12 Multi-Step Problem and Rubric 109
4.12.1 The Multi-Step Problem 109
4.12.2 The Rubric 111
4.13 End-of-Semester Survey 113

5 COMPARISON OF SCAFFOLDINGS 115
5.1 Preface 115
5.2 Introduction 115
5.3 Background 117
5.3.1 Description of the Class 117
5.3.2 Scaffolding and Problem Solving 118
5.3.3 Scaffolding in CITA 120
5.4 Research Design 123
5.5 Research Participants 124
5.6 Quantitative Methods 124
5.7 Qualitative Methods 126
5.8 Results 128
5.8.1 Overall Use of Immersive CITA 128
5.8.2 Overall Use of Postscripts 130
5.8.3 The Truncated Step-by-Step Structure vs. the Detailed Step-by-Step Structure 131
5.8.4 The Branching Structure 133
5.8.5 Student Retention as They Traverse a Tutorial 135
5.9 Discussion 137
5.10 Conclusions 139
5.11 Acknowledgments 140

6 SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS 141
6.1 Summary 141
6.2 Research Questions 141
6.3 Future Directions 143
6.3.1 Longitudinal Study of CITA 144

Page

6.3.2 Create Multiple Paths with Different Tutorials 144
6.3.3 Implement "Dead End" Style Tutorials 145
6.3.4 Conduct a Detailed Graphical User Interface Study 145
6.4 Implications of the Research 146

LIST OF REFERENCES 147

A CHEATING ON THE BEMA CONCEPT INVENTORY 157
A.1 Preface 157
A.2 Introduction 157
A.3 Background 159
A.3.1 Overview of the Classes 159
A.3.2 Class Demographics 159
A.3.3 Administration of the BEMA 162
A.4 Methods of Analysis 163
A.4.1 Measures of Reliability and Discrimination 163
A.4.2 Roles of the Researchers 165
A.5 Results 165
A.5.1 Classical Test Theory 165
A.5.2 External Validity 178
A.6 Discussion 179
A.6.1 General Trends 179
A.6.2 The Outlier 181
A.7 Conclusions 183
A.8 Acknowledgments 183

B TRANSCRIPTS 185

VITA 186

LIST OF TABLES

Table Page

1.1 Enrollment in the campus and online sections of Electricity and Optics from the spring of 2014 through the summer of 2016. 3

3.1 The schedule of development of the CITA system. In the spring semester of 2015 and prior, the CHIP homework system included 29 "interactive examples" that were originally developed by the faculty at UIUC (out of 139 homework problems total). These 29 problems were paired with step-by-step tutorials and made up approximately 21% of the homework problems seen by students. 49

4.1 Enrollment in the campus and online sections of Electricity and Optics from spring 2014 through summer 2016. All of the enrollment numbers ignore dropped students. 60

4.2 The schedule of development of the CITA system. In the spring semester of 2015 and prior, the CHIP homework system included 29 "interactive examples" that were originally developed by the faculty at UIUC (out of 139 homework problems total). These 29 problems were paired with step-by-step tutorials and made up approximately 21% of the homework problems seen by students. 68

4.3 P-values of the ANOVA analysis of the number of clicks between males and females in different sections and semesters. 77

4.4 P-values of the chi-squared analysis of the number of clicks between males and females in different sections and semesters. 77

4.5 Exit survey results of students in PHYS 24100, broken down by semester and PHYS 17200 grade. The question stated "I used the interactive walk-throughs to help me with homework problems". 79

4.6 Chi-squared analysis of scores on selected test questions between the on-campus (top) and online (bottom) sections of the fall 2015 (fa15) and spring 2016 (sp16) semesters. We compare correct (C) and incorrect (I) answers between the semesters. The highlighted rows show statistically significant improvements between the fa15 and sp16. 89

4.7 Chi-squared analysis of selected test questions between pre-CITA and CITA semesters of PHYS 24100. There are no statistically significant differences between the semesters (save one question). 91

Table Page

4.8 BEMA gains versus reported use of the CITA system for each semester. Spring 2016 shows the first statistically significant difference in gains between CITA users and non-users. 96

4.9 Scores on the multi-step problem in each semester when it was administered, along with the average score and standard deviation of the semester. 98

4.10 Scores on the multi-step problem for each semester were divided into groups by responses on question one of the survey. A chi-squared test was used to determine if the difference in the distribution of reported uses was statistically significant. 98

4.11 Rubric for the analysis of the multi-step problem. In addition to these six possible points, graders made a note of whether a student drew a diagram within their solution and whether a student tried to apply the time rate of change of momentum in the solution (F = dp/dt). 112

5.1 Enrollment in the campus and online sections of Electricity and Optics from spring 2014 through summer 2016. 118

5.2 Fraction of students that start an Immersive CITA tutorial, averaged over the semester and section. 129

5.3 The fraction of students who attempted the first of the follow-up questions in the postscript, averaged over the 27 test problems. 130

A.1 Enrollment in the campus and online sections of Electricity and Optics from spring 2015 through spring 2016. All of the enrollment numbers ignore dropped students. 160

A.2 Reported mean mechanics grades for students in PHYS 24100. Students were asked to report their grade in their previous mechanics course (with the option of not disclosing this information should they want to keep it private). The majority of students took PHYS 17200 - the introductory mechanics course at Purdue University. The options were “A” (4), “B” (3), “C” (2), and “D” (1). 161

A.3 Reported mean calculus grades for students in PHYS 24100. Students were asked to report their grade in their most recent calculus course (with the option of not disclosing this information should they want to keep it private). Most students had completed the second semester of introductory calculus and were taking Calculus III concurrently with PHYS 24100 or 24100D. The options were “A” (4), “B” (3), “C” (2), and “D” (1). . 161

Table Page

A.4 The p-values of the chi-squared tests comparing scores in the online and on-campus sections. We only compare grades of A, B, and C since the number statistics were very low for D grades. Wherever there was a statistically significant difference in the distributions, we saw a small effect size. . . 162

A.5 Average item difficulties on the BEMA. The data below is for the spring 2015 (SP15), summer 2015 (SU15), fall 2015 (FA15), and spring 2016 (SP16) semesters. 168

A.6 Average point biserial correlation coefficients for the BEMA pre-test and post-test in each semester. There is a distinct difference between the online and on-campus sections of Electricity and Optics. 173

A.7 Ferguson's Delta for the BEMA pre-test and post-test for each section and semester. 175

A.8 KR-21 reliability index for the BEMA pre-test and post-test for each section and semester. 177

LIST OF FIGURES

Figure | Page

1.1 Enrollment in Electricity and Optics in the spring and summer semesters. Rising summer enrollments are reducing the class sizes in the spring semesters. 4

1.2 A screen shot of a homework problem on the CHIP homework system. Students are required to answer the "top-level" question correctly for credit (as shown above). However, most questions also included optional tutorials and follow-up exercises that would guide students towards the correct answer. 5

3.1 A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error. 45

3.2 A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion. 46

3.3 A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer. 47

4.1 A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error. 65

4.2 A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion. 66

4.3 A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer. 67

4.4 The growth in the number of clicks between the different semesters of PHYS 24100. The x-axes give the number of clicks while the y-axes give the number of students. There were 138 homework questions given in each semester. 76

Figure Page

4.5 Overall course grade as a function of number of clicks within the CITA tutorial system. The students are binned by their previous grade in PHYS 17200 (i.e. introductory mechanics). 83

4.6 Overall course grade as a function of number of clicks within the CITA tutorial system. The students are binned by their previous grade in introductory calculus (i.e. Calculus 1). 84

4.7 Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by gender. . 85

4.8 Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by domestic and foreign status. 86

4.9 Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by their section (i.e. online vs. on-campus). 87

4.10 Overall course grade in PHYS 24100 compared with student answers on the exit survey. 88

4.11 Total exam grade as a function of number of clicks within CITA. 93

4.12 Student gains on the BEMA as a function of reported CITA use. The students shown in these plots are all in on-campus sections of PHYS 24100. 94

4.13 Student gains on the BEMA as a function of reported CITA use. The plot on the left shows the unfiltered online gains in the summer of 2015. The plot on the right filters out potential cheaters, thus giving a more accurate representation of the system. 95

4.14 Student gains on the BEMA as a function of reported CITA use in the campus section of spring 2016. 97

4.15 A comparison of the maximum number of nodes taken by a student within a CITA tutorial on a question-by-question basis. Each point in the graph represents a single question on the homework, while the x- and y-axes represent the maximum number of nodes visited by a student in the online and on-campus sections of PHYS 24100. The blue line in the diagram marks the spot where online and on-campus sections use the tutorials equally. 100

5.1 A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error. 121

Figure Page

5.2 A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion. 122

5.3 A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer. 123

5.4 Examples of directed graphs generated using Matlab. Each of the points in the graphs refers to a specific step in an Immersive CITA tutorial or the graded question in the homework. The lines in the graph are weighted and represent the number of students that move from step to step in a problem. 125

5.5 Histograms showing the use of CITA in the fall 2015 semester (left) and the spring 2016 semester (right). The horizontal axis shows the fraction of students who used an Immersive CITA tutorial while the vertical axis displays the number of problems. 129

5.6 A comparison of the choices that students made when traversing the branching problems. The left graph shows the on-campus sections while the right graph shows the online sections. The number of students who traversed a given path in the branching problem was normalized with respect to the total number of students in the class. 134

5.7 Histogram of the fraction of students who work the general path through to the end. The median is plotted as a blue line. 136

5.8 Histogram of the fraction of students who work the general path through to at least one step before the end of the tutorial. The median is plotted as a blue line. 136

5.9 Fraction of students who completed each step in an Immersive CITA tutorial. There is a sudden drop in retention at certain steps of the tutorial. 137

A.1 Distribution of BEMA scores for the online and on-campus sections of Electricity and Optics. The red bars represent the pre-test scores while the blue bars represent the post-test scores. The on-campus sections follow a normal distribution with a slight skew (as one might expect). However, the online sections have an abnormal peak at a score of 30/31. Note that there were no on-campus sections in the summer of 2015. 164

A.2 Item difficulties for online and on-campus students of PHYS 24100. Each point refers to a question on the BEMA. 166

Figure Page

A.3 Comparison of pre-test and post-test item difficulties for the online and on-campus sections. The graph on the left shows the data in the fall 2015 semester while the graph on the right shows the data from the spring 2016 semester. 167

A.4 The figure shows the item discrimination indices for the online and on-campus sections of the spring 2015 through spring 2016 semesters. Each point on the graph represents a question on the BEMA. In the spring 2015 graph, one of the questions had an index less than zero, so it does not appear on the graph. It is also important to note that the grid-like patterns of the online sections are an artifact of lower than ideal number statistics. Thus, they should not be interpreted as an underlying pattern in the data. 170

A.5 The figure shows the pre-test and post-test item discrimination indices for the spring 2015 through spring 2016 semesters. Each point on the graph represents a question on the BEMA. In the spring 2015 graph, one of the questions had an index less than zero, so it does not appear on the graph. The lines with slope one are plotted for the sake of comparison. 171

A.6 The figure shows the point biserial correlation coefficient for the online and on-campus sections of the fall 2015 semester. The clustering of the data points shows that students are not experiencing the same test in the online and on-campus sections of Electricity and Optics. 172

A.7 A comparison of exam scores compared with the BEMA pre-test (red) and post-test (blue) scores in the fall 2015 and spring 2016 semesters. The fall 2015 semester is the top row of graphs while the spring 2016 semester is the bottom row of graphs. The graphs on the left show the data from the campus sections while the graphs on the right show the data from the online sections. Each point represents the total test score of one student on the first two exams (i.e. material that is covered on the BEMA). Thus, students who took both the BEMA pre-test and the post-test appear twice in each graph (i.e. one red data point and one blue data point). The online students who are identified as cheating are found on the right hand side of the graphs, corresponding to high BEMA scores but average exam scores. 180

ABBREVIATIONS

C3PO	Customizable Computer Coaches for Physics Online
CAPA	Computer Assisted Personalized Approach
CHIP	Computerized Homework in Physics
CI	Concept Inventory
CITA	Computerized Interactive Teaching Assistant
CMS	Course Management System
E&M	Electricity and Magnetism
GUI	Graphical User Interface
IE	Interactive Engagement
ITS& Intelligent Tutoring System	
JITT	Just-in-Time Teaching
LON	Learning Online Network
MOOC	Massive Open Online Course
OWL	Online Web Learning
PER	Physics Education Research
SPOC	Small Private Online Course
STEM	Science, Technology, Engineering, and Mathematics
ZPD	Zone of Proximal Development

GLOSSARY

constructivism	a learning philosophy that states that people construct and revise their understanding of the world through active experience, not passive absorption of ideas
distance learning	a pedagogical approach where educational institutions provide learning resources to students who are located far from the school
epistemic	relating to the description or validation of knowledge
metacognition	the analysis of one's own thought process (e.g. planning how to approach a given task, monitoring one's own comprehension, and evaluating progress toward the completion)
ontology	the philosophical study of the nature of existence
self-concept	a set of self-constructed beliefs that one holds about oneself
self-efficacy	one's belief in his or her ability to accomplish a given task

ABSTRACT

Vandrevala, Cyrus M. Ph.D., Purdue University, December 2016. The Development of a Computerized Interactive Teaching Assistant in Physics: The CITA on CHIP Project. Major Professors: Lynn Bryan and Andrew Hirsch.

One of the many roles of university instructors is to provide help to students throughout the semester - especially in the form of feedback on homework. Personalized feedback from the instructor might be possible in a small classroom setting, but becomes unmanageable when class sizes grow to dozens or even hundreds of students. As a result, universities are turning to computerized homework systems that guide students through problems and provide focused grades and feedback.

The overarching goal of this project was to design a comprehensive, scaffolded set of tutorials for the homework questions in an undergraduate electricity and magnetism course for engineering majors and determine how the tutorials affected student performance. The research group developed CITA (Computerized Interactive Teaching Assistant) using a design-based research approach. This system assessed student knowledge and provided focused feedback using three structures - Shallow CITA, Immersive CITA, and Postscripts. Throughout the semester, I collected student grade data as well as student opinions of the CITA system through surveys and focus group interviews. The analysis of data informed an iterative development process.

This study aimed to answer a few research questions. First, does the use of CITA tutorials improve student performance in their introductory electricity and magnetism course? Second, what are students' views of the CITA program? Finally, what motivates students to choose CITA over some other external educational resource?

Overall, the first version of CITA showed no gains in student performance in the summer of 2015. However, versions two and three of CITA showed small but

statistically significant gains for students that used the system. This is especially apparent in the spring 2016 semester.

Students who used versions two and three of CITA generally had positive things to say about the system. Surveys questions based on Likert scales along with open-response questions and focus group interviews all indicate that students found Shallow CITA particularly helpful and Immersive CITA helpful under most circumstances. However, students generally did not find the Postscripts helpful or worthwhile.

Finally, I conducted a step-by-step analysis of the tutorials that were used in the CITA system in order to track student retention. These results were synthesized with student responses from focus group interviews to gain an understanding of what motivates students to use CITA tutorials over other external resources (e.g. Google, Yahoo Answers, the course textbook, etc.). Since students were able to choose between CITA and other external resources, they adopted strategies to efficiently solve homework problems rather than maximize learning. Students left a CITA tutorial if they encountered a step which was perceived as tedious. Thus, online tutorials must be designed to prevent students from prematurely leaving the tutorial system.

1. INTRODUCTION

1.1 Background

Due to the increased popularity of online classes and the rising enrollment in foundational, university-level science courses, professors are under constant pressure to provide large numbers of students with assistance throughout the semester. Weekly homework assignments, in particular, require instructors to dedicate a large amount of time and energy to help students with their questions. When direct assistance from a faculty member is not available, students need a learning tool that guides them through homework problems. This is especially important in an environment where non-traditional students with diverse backgrounds must attend class [1–3]. Classroom communication systems (i.e. classroom response systems) are one type of tool that have been utilized with great success during lectures [4, 5]. However, the expectation of access to faculty at any time of the day outside of class is unrealistic; such is the case for the calculus-based, introductory physics class at Purdue University - a course with an enrollment of over 700 students during the fall semesters.

The Computerized Interactive Teaching Assistant (CITA) was conceived to guide students through homework problems while teaching the nuanced techniques of physics problem solving. Traditional computerized homework systems typically provide a “correct” or “incorrect” response to a given answer with only a general explanation as to why their logic was wrong. Often, these systems do not provide specific instructions on how to proceed from the point at which the student made the mistake. Instead, there may be a link to a passage in the textbook that outlines a derivation without any specific context. If this level of assessment is not helpful, the student may either contact an instructor, consult with a peer (who may be equally confused), or wait for recitation later in the week. Thus, the learning process is interrupted as the student waits for simple guidance. Conversely, an intelligent tutoring system aims to provide focused feedback to students when it is required [6–8]. The goal of the CITA project is to design, implement, and analyze an online system that is readily available to provide focused guidance to help students learn physics concepts and solve homework problems.

1.2 Overview of Electricity and Optics

The CITA project took place in the undergraduate, calculus-based, introductory E&M courses at Purdue University: PHYS 24100 and PHYS 24100D. These courses are developed for sophomore engineering majors, with the exception of the electrical engineering majors who take PHYS 27200 [9]. Both PHYS 24100 and PHYS 24100D are titled "Electricity and Optics"; the "D" simply signifies a distance learning (i.e. online) section of the course in the fall and spring semesters. The content in the on-campus and online sections is exactly the same, and all of the students use *Physics for Scientists and Engineers* by Tipler and Mosca as their textbook [10]. They only differ in the fact that the distance learning sections watch video lectures and attend online recitations through Cisco WebEx [11], while the regular sections watch live lectures and attend live recitations on campus.

Engineering students usually take this course in the fall semester (the "on-semester"). Alternatively, some engineering students choose to take the course in the spring or summer (the "off-semesters") due to the fact that they are either ahead or behind in their plan of study. Table 1.1 shows the total enrollment in the campus and online sections of Electricity and Optics over the last three years. During each of the fall and spring semesters, just under 10% of the total class is enrolled in PHYS 24100D. However, a new trend emerged in the spring of 2016 - more students took PHYS 24100 in the summer rather than in the spring due to the rising popularity of the online class (see Figure 1.2).

Table 1.1
Enrollment in the campus and online sections of Electricity and Optics from the spring of 2014 through the summer of 2016.

Semester	**Online**	**Campus**	**Percent Online**
Spring 2014	41	465	8.1%
Summer 2014	153	0	100.0%
Fall 2014	62	725	7.9%
Spring 2015	43	474	8.3%
Summer 2015	187	0	100.0%
Fall 2015	69	728	8.7%
Spring 2016	36	339	9.6%
Summer 2016	222	0	100.0%

Electricity and Optics is generally taken after students complete PHYS 17200 - Modern Mechanics. The introductory mechanics course is currently taught using the first semester of the *Matter and Interactions* curriculum by Ruth Chabay and Bruce Sherwood [12]. Some of the major topics covered in Electricity and Optics are electric charge, electromagnetic fields, Maxwell's equations, geometric optics, and interference effects. Due to the diverse nature of topics in PHYS 24100 and PHYS 24100D, this course is a prerequisite for many of the intermediate engineering courses at Purdue University.

1.3 The CHIP Homework System

PHYS 24100 and PHYS 24100D use an in-house homework system called Computerized Homework in Physics (CHIP). The CHIP system is based on the CPlite homework system from the University of Illinois at Urbana-Champaign (UIUC). Purdue University adopted an early version of CPlite in 1997 and continued to make

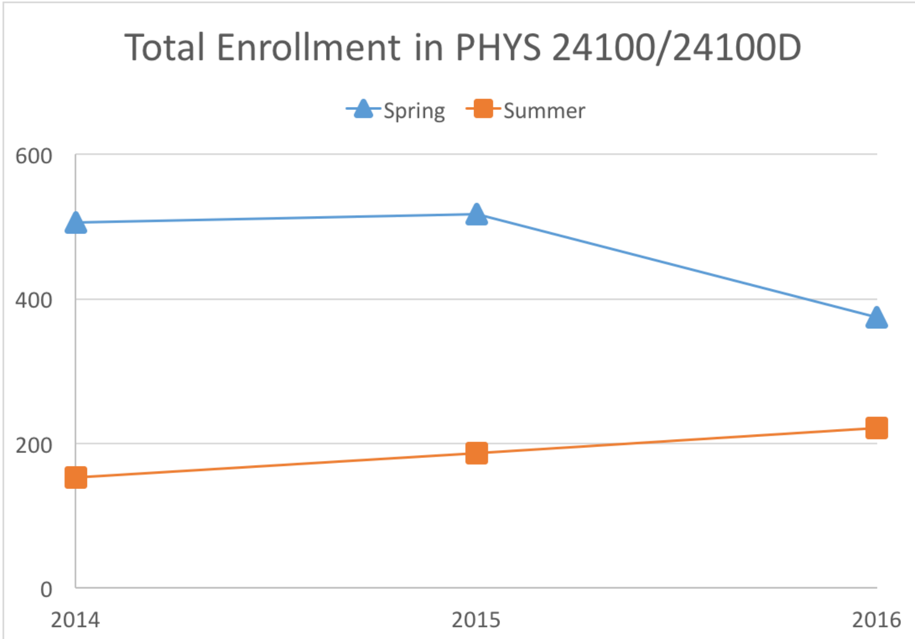

Figure 1.1. Enrollment in Electricity and Optics in the spring and summer semesters. Rising summer enrollments are reducing the class sizes in the spring semesters.

improvements to it over the next decade, adding statistical analysis tools, grade books, and updated homework problems. CHIP serves not only the introductory physics courses but a few other physics courses at Purdue University [13, 14]. Figure 1.3 shows a screen shot of a sample problem on CHIP.

CHIP is a well-established system that has over a decade of in-the-field use. Even so, at the beginning of the research project, the research group had to decide if it was better to develop an entirely new homework system or update CHIP to suit our research purposes. We decided to update the existing CHIP system for a variety of reasons.

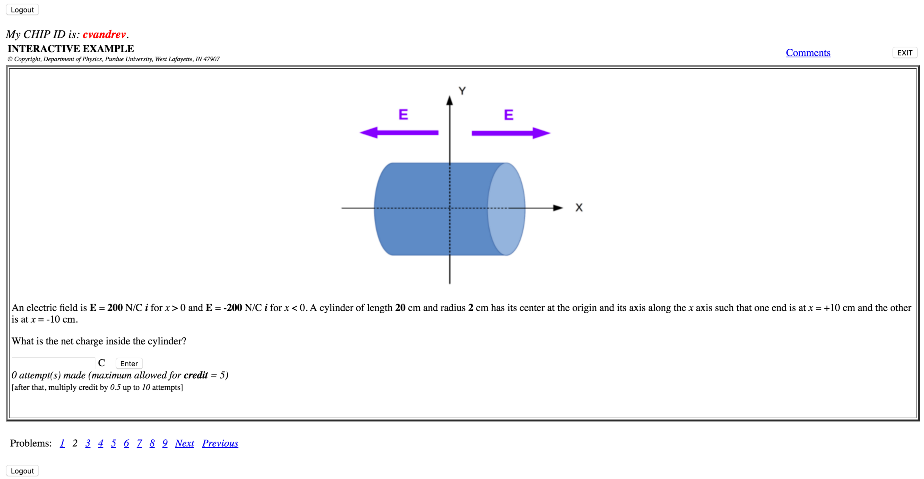

Figure 1.2. A screen shot of a homework problem on the CHIP homework system. Students are required to answer the "top-level" question correctly for credit (as shown above). However, most questions also included optional tutorials and follow-up exercises that would guide students towards the correct answer.

1.3.1 Strengths of CHIP

CHIP has a well-developed database of problems from a variety of textbooks including: Tippler and Mosca [15]; Giambattista [16]; Halliday, Resnick, and Walker [17]; and Cutnell [18]. These problems are free to use and modify, as long as the changes are contained within the CHIP system. Thus, the physics faculty members at Purdue have ready access to hundreds of potential homework, quiz, and exam questions. Instructors in the physics department are free to recommend updates to the system and create new questions. Most importantly, it is possible to undertake the development of these new questions without interrupting any existing classes.

All of the CHIP servers are maintained on Purdue University's West Lafayette campus in the physics building by the university's support team. CHIP has a dedicated staff of experts that run and maintain the system; this includes technology professionals who maintain the stability and security of the system and faculty members who use it to assign and grade homework. No aspect of CHIP has to be outsourced to a third party. One advantage of this is when students ask questions or send error reports, they are talking directly with a course administrator rather than a secondary source. Thus, questions and errors can be resolved very efficiently.

Finally, CHIP includes a grade book that can collect student scores, filter the data by demographic information, and perform basic statistics between groups of students. If CHIP is not capable of performing a specific statistical technique, the data can be formatted and exported to CSV files which can be read into a program like R or MATLAB. Since CHIP keeps a secure archive of scores from the past, one may easily compare students within a semester and between semesters.

1.3.2 Weaknesses of CHIP

Although CHIP has a successful history at Purdue University, it is not without its problems. The CHIP program was created just as the Internet was beginning to receive mainstream acceptance. Thus, while the graphical user interface (GUI) was revolutionary at one time, it has begun to show its age. The overall look and layout of CHIP has not kept up with modern devices. Specifically, buttons and menus do not render well on small screens like cell phones, tablets, and net books.

CHIP has a lot of underutilized features for analyzing student performance. Although many statistical analysis tools are available on the system, they are rarely used due to the fact that they are either hard to find or have a steep learning curve (i.e. Perl programming skills might be required). This means that professors often perform their analysis in other places, even if the software is right in front of them.

Although CHIP has resources for building tutorials within individual problems, at the start of this research project only 21% of the assigned homework problems had any sort of tutorial. These original tutorials, also called “interactive examples”, were developed by the faculty at UIUC, and were utilized at Purdue University. They offered help only in a step-by-step fashion. Thus, the homework problems were in need of an update.

1.4 The Need for CITA

There was a pressing need for the development of CITA in PHYS 24100 at Purdue University. In addition to the weaknesses of CHIP, student comments in CHIP problem reports and course evaluations supported the need for an interactive tutorial system.

Many students commented that the homework system did not provide good feedback in the cases when their work was correct but contained simple mistakes like unit errors. Nathan P. commented that his overall experience with CHIP was soured because of a simple mistake:

> With all due respect, I am very displeased with CHIP. It took me 45 minutes to solve one problem because I didn't round correctly. I know it is not my place to say this, but I think that online HW submission programs should be done away with at universities like Purdue. With that being said, I am going to power through this year in physics! (Nathan P., Fall 2014 Semester)

Although we may disagree with completely doing away with online homework systems in large universities like Purdue, this problem is very valid; the homework system did need some way to steer students in a productive direction, especially if they were only making a simple rounding error. Many other students have commented on the need for some sort of tool to help them get through the homework in their course evaluations.

As of the start of this project, some of the homework problems included rudimentary tutorials. These interactive examples made up a small percentage of the total number of problems. However, they were very popular among students. Every semester, students commented on how helpful these examples were:

> Thank you for the help section in this problem! It was very well written and helped me figure out the problem and (hopefully) similar ones in the future! (Walt R., CHIP Error Report from Fall 2014)

> This [tutorial] is really helpful. Understanding how to set up the integrals has always been a challenge, and this makes it clear how to break it apart according to the electric fields. (John C., CHIP Error Report from Fall 2014)

> I found the IE [interactive examples] extremely helpful, and want these to continue to be available. It was much more helpful than examples in the book or just completing the homework, and I think it'd be great for more complicated problems especially. (Shaun A., CHIP Error Report from Spring 2015)

However, many of the existing interactive examples were in desperate need of a revision. This student comments on an existing tutorial in a geometric optics problem:

> It was confusing to start by entering theta 5 and working backwards rather than starting at theta in and working forwards. If the student knows what theta 5 is already, they probably wouldn't need to go through the help for that part. (Nicholas L., CHIP Error Report from Spring 2015)

These examples were representative of the need for a new tutorial system. The interactive examples that we developed in this project were positively received by the student body and can be expanded to assist future PHYS 24100 students. "We" refers to the team of researchers that developed, implemented, and analyzed CITA in PHYS 24100 / 24100D. The development and implementation of CITA on CHIP required work on many fronts:

1. We designed a framework that can be used to build online tutorials with multiple paths of analysis. This was completed largely due to the effort of Dr. Hisao Nakanishi.

2. We converted traditional homework problems into interactive examples. Each of these interactive examples include hints and detailed physics tutorials.

3. We updated some of the outdated web pages on CHIP to make the menus more accessible on desktop computers, laptops, tablets, and cell phones.

4. We tracked student progress throughout the semester as a function of their use of CITA. We then used this data to implement continuous improvements from semester to semester.

5. We aligned the teaching methods used in the classroom with those used on CITA. This is certainly an area of continuing development.

It is not enough to simply develop a tutorial system without an organized set of research questions that can guide the development and analysis. These will be discussed in the next section.

1.5 Research Statement

The National Science Foundation has stated that solving complex, real-world problems and creating structured experiments are the most important knowledge and skills that students should develop in their educational careers [19]. However, science, technology, engineering, and math (STEM) students often do not have these skills and knowledge when they start their college careers or even after they earn their degree. Thus, universities must provide the opportunities and support to help students become accomplished problem solvers.

It is not enough just to develop tutorials; a homework system needs to be evaluated in order to determine how it promotes student learning. I was specifically interested in analyzing how the structure of the CITA system affected student problem solving. Instead of relying on a linear sequence of steps to reach an answer, the CITA system allows students to traverse different paths with different tutorials as they make their way to the answer. I wanted to examine how this unique type of scaffolding influenced the way students solve problems in physics. Thus, the overarching questions for this research study follow below:

How does the branching structure of the interactive examples in CITA influence student problem solving abilities?

Do students develop problem solving skills as they use the CITA system? For example, are they able to think through derivations and non-trivial multi-step problems after the scaffolding is removed? Do students depend on resources like the textbook to guide their analysis rather than starting from first principles? How does CITA enhance problem solving skills? Research findings regarding this question will be addressed in Chapter 4.

What are students perceptions about the CITA system?

How do students like the CITA system, and how does it compare with other homework systems that exist today? Research findings regarding this question will be addressed in Chapter 4.

What motivates a student to use CITA rather than another external source?

Which students are using the CITA system? Are there distinct groups that use the system more than others? Additionally, are there any specific parts of the tutorials that cause students to turn to other external resources? Research findings regarding this question will be covered briefly in Chapter 4 and in more detail in Chapter 5.

1.6 Organization of the Thesis

Due to the number of papers, posters, and proceedings that have stemmed from this research project, this thesis will make use of external publications to highlight major results. Each chapter of the thesis will focus on a specific topic, including a paper where appropriate. Chapters that incorporate a pre-print or published paper will have an introductory section that will set up the context for the paper and results. Finally, the chapters have been organized to tell the narrative of the CITA research and development project.

Chapter 2: Theoretical Framework and Literature Review

The literature review provides an overview of the learning theory of constructivism and the differences between master and novice problem solvers. I describe how scaffolding can help students solve physics problems and review some of the online homework systems that are in use today.

Chapter 3: Methods

Since many of the chapters include a paper that will be submitted for publication, the specific details of the methodology are dispersed throughout the thesis. Thus, I have collected and summarized the details of the methodology in a single chapter so the reader can easily identify my research methods.

Chapter 4: An Overview of CITA

This chapter provides an analysis of the CITA system. Specifically, I describe the development process, the theoretical framework, and the analysis techniques. Then, I provide a full report of how the tutorials influenced student problem solving skills.

Chapter 5: Node Analysis of the CITA Tutorials

In this chapter, I provide a second analysis of the CITA system through a different lens. First, I describe how the individual tutorials were designed and structured. Then, I present an analysis of how students traversed different tutorial structures over different semesters. Finally, I discuss student strategies and motivations while using CITA tutorials.

Chapter 6: Summary, Implications, and Future Directions

This final chapter provides a brief overview of the previous chapters and describes how the results are related. I then outline the major implications of the study and the potential future directions of development.

Appendices

The appendices contain the supplementary materials that were used in the study (i.e. surveys, quizzes, etc.). Additionally, I include a research paper that analyzes cheating between online and campus sections in the BEMA exam.

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

> Education research is the scientific field of study that examines education and learning processes and the human attributes, interactions, organizations, and institutions that shape educational outcomes. Scholarship in the field seeks to describe, understand, and explain how learning takes place throughout a persons life and how formal and informal contexts of education affect all forms of learning. Education research embraces the full spectrum of rigorous methods appropriate to the questions being asked and also drives the development of new tools and methods. (American Educational Research Association [20])

2.1 Background

Due to the fact that learning hinges on so many factors, it is not enough to be an expert in one discipline while ignoring all others. Education researchers need a diverse, yet complementary, knowledge of many areas: learning theories, teaching strategies, motivation, psychological testing, and classroom management just to name a few. Furthermore, these domains are further refined when studying different populations - students, teachers, administrators, etc.

In this chapter I present the theoretical basis for the development of the online homework system and this research study. I first summarize the learning theory of constructivism and how it relates to this project. Next, I describe how scaffolding methods enhance student problem solving skills. Then, I discuss how multimedia learning has shaped the development of the CITA tutorials. Finally, I describe some of the current homework systems on the market and future directions of physics education research within the context of online homework systems.

2.2 Constructivism

A learning theory (or learning philosophy) describes how people capture, process, retain, and interpret information. Many learning theories have been put forth throughout history. Some are very simple while others include cognitive, emotional, environmental, and social effects on the learner. Constructivism is an overarching term for a learning theory (and its many variants) that date back to the 1800s; it has seen a rise in popularity over the past few decades due to the work of researchers like vonGlasersfeld [21] and Bandura [22] and will be the starting point for this discussion.

When a student learns something new, he or she has some prior knowledge of or related to the subject. This knowledge can come from formal study (such as a previous physics class) or from life experiences (seeing a ball rise into the air and fall back down to the ground after it has been thrown). Prior knowledge includes both normative models of how the world works as well as non-normative, pre-existing notions of the universe. No matter if these assumptions are correct or incorrect, they stem from the fact that students notice regularities in the world that are confirmed by repeated experience [21, 22]. A constructivist would posit that learning is an active process where new information about the world is linked to a learner's previous models of how the world works.

That being said, constructivists hold varying perspectives on the mechanisms through which this linking takes place. Their ideas fall along a spectrum that includes radical constructivism, cognitive constructivism, and social constructivism. Radical constructivism, an idea popularlized by Ernst vonGlasersfeld, says that each individual constructs a "personal reality". It can be summarized by two major claims [23,24]:

1. Knowledge is not something that is passively received. Instead, it is actively constructed, generated, and revised by a person over time as he or she fits new

information about the world to his or her pre-existing notions of how the world works.

2. This adaptive knowledge is meant to organize the world. An organized worldview does not necessarily mean a correct worldview.

Thus, radical constructivism rejects the claim that the discovery of an ontological reality is the end goal of learning. Rather, it claims that the end goal of learning is organizing the world. Learning is subjective and can be skewed by personal bias, faulty senses, and previous experiences.

This idea is especially important when teachers plan lessons. A radical constructivist would argue that learning is the process of adjusting models with each new experience. Due to the fact that students build models rather than passively receive knowledge, there will be a "radical separation" of results between educational procedures that strive to generate a fundamental understanding of the material and those that encourage memorization. In order to be an effective teacher, one must not only teach the material, but infer the thought processes taken by students. This means that the instructor must be especially interested in the errors made by students, for those errors are the points at which true learning can begin [25].

Social constructivism stresses that one's environment is vitally important to learning. No man is an island; a person uses his or her cultural and social setting in developing a model of how the world works. This reality combines a personal interpretation of the universe with an interpretation shared by one's group of contacts. Social constructivists agree with radical constructivists that a person's reality is actively constructed. However, social constructivists would state that without a social network, there can be no learning. In other words, knowledge is a product of social interactions.

Social constructivism is rooted in the works of Piaget [26–28], Vygotsky [29, 30], and Bruner [31–34]. Albert Bandura has also indirectly influenced social constructivism through his work in social-cognitive theory [35–37]. However, many would consider him more aligned with the behaviorist learning philosophy than the constructivist philosophy. Nonetheless, Albert Bandura succinctly summarized the philosophy of social learning in the following quote:

> Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do. Fortunately, most human behavior is learned observationally through modeling: from observing others one forms an idea of how new behaviors are performed, and on later occasions this coded information serves as a guide for action. (Social Learning Theory, p. 22, [38])

Cognitive constructivism also asserts that the acquisition of knowledge is an adaptive process that results from active model building by an individual learner. However, it is not as extreme as radical constructivism in that it posits that there is an independent reality that is knowable to the individual. Therefore, learning is the result of adjusting mental models to more accurately reflect external reality. Unlike social and radical constructivists, a cognitive constructivist would claim that an objective reality does exist, and learners can probe it directly.

In this study, I use the learning theory of cognitive constructivism to guide my development and analysis. The CITA tutorials are designed for individual students to complete at their own discretion. Although students can certainly discuss the content of the tutorials with their colleagues, social interaction is not a pre-requisite for their use. Additionally, the instructors of Electricity and Optics as well as the research team make the assumption that classical electromagnetism does accurately describe nature within a specific context. Thus, cognitive constructivism aligns with the assumptions of the research team.

In the following sections, I will describe how cognitive constructivism is applied to problem solving in physics.

2.3 Polya's Problem Solving Strategy

The modern theory of problem solving is rooted in the work of George Polya [39]. He outlined a deceptively simple strategy for solving a problem:

1. Understand the Problem
2. Devise a Plan
3. Execute the Plan
4. Review the Work

Although this might seem completely trivial, many problem-solvers do not make it past the first step [40]. I summarize Polya's descriptions of the four steps below and then explore them in the context of physics education research.

2.3.1 Understand the Problem

Understanding the problem is more than just knowing the variables that one is given. One must take the time to fully understand the domain of what is being asked. For example, suppose a physics student is given a word problem. She will probably correctly identify what she needs to solve for and what she is given. However, there are many aspects of the problem that go unanalyzed. Is there enough information to solve the problem? What assumptions need to be made in order to solve the problem? Can she restate the problem in a simpler way? What is the required tolerance for the answer? All of these details are important to obtain a correct solution to the problem.

2.3.2 Devise a Plan

There are a number of strategies that can be used to solve a given problem. On one extreme, a student might use the guess-and-check strategy to try to find the answer. On the other extreme, a student might work backwards from a known model to try to identify similarities between what he knows and what he needs to find. Polya asserts that the student needs to take the time to figure out what the best strategy to solving the problem might be. More often than not, if the student does not have a well-defined plan to solve a problem, then he will turn to an external source for a plan (e.g. textbook, class notes, internet).

2.3.3 Execute the Plan

Executing the plan does not just mean plugging numbers into equations. It is a process by which a student checks her work to make sure that it is in line with her understanding of the problem and the plan that she devised. The physics student needs to carefully comb through the solution, checking her work for mathematical or conceptual errors. Additionally, the final step to the problem solving process should be assessing the answer to see if it makes sense within the context of their knowledge of physics.

2.3.4 Review the Solution

The analysis of a problem does not stop once a solution is found. That solution needs to be evaluated in the context of the problem. For example, does the solution make sense? Does it coincide with all of the assumptions made during the analysis? Are there ways that the solution can be improved in the future? This final step is just as important as the ones before it.

2.4 Problem Solving in Physics

2.4.1 Novice Problem Solvers

Because of the importance of problem solving in our society, there has been a great deal of research that investigate how students learn to solve problems in physics.

Many of these studies make the distinction between “master” and “novice” problem solvers. A novice problem solver is one who is inexperienced in the art of solving complicated problems. He or she generally does not execute a systematic method such as Polya’s method described above, but instead works in a potentially erratic way [41]. Most students at the start of their undergraduate careers would be classified as novice problem solvers.

Walsh and her team conducted a phenomenographic study of students problem solving approaches in introductory physics. She interviewed a series of students in order to determine their specific approaches to solving physics problems. She then categorized them into different approaches including a scientific approach, a plug-and-chug approach, a memorization approach, or no approach. She confirmed that the majority of young, undergraduate students do not start off with sophisticated problem solving skills and approach problems in a non-scientific manner [42].

Redish et al. identified a number of “epistemic games” that students play to work through problems in physics [43]. For example, they describe how novice problem solvers often work backwards in order to solve physics problems - skipping the initial analysis described by Polya and instead fitting equations to known quantities until a pattern is uncovered.

Sweller’s summary of “means-end” analysis techniques supports other authors mentioned in this section. “Means-end” techniques describe how students work backward from a final goal by setting sub-goals. This continues until an equation with no further unknowns is encountered. Sweller found that this method of analysis imposes a heavy cognitive load on the student and does not help the student build an understanding of the problem. He further stated that a curriculum that is heavily focused on problem solving may not aid a student in schema acquisition (i.e. a knowledge structure based on fundamental principles) [44].

Overall, researchers have found that novice problem solvers adopt strategies that identify surface features in problems. For example, they might relate two problems due to the fact that they have similar diagrams without realizing that the underlying concepts that govern the analysis in each problem are very different. This is very different than expert problem solvers.

2.4.2 Expert Problem Solvers

On the other hand, expert problem solvers follow a structured problem solving strategy. They start by characterizing the qualitative aspects of a problem before developing a mathematical solution. They are not distracted by similar looking diagrams or solutions stated in the book. They start from first principles and work their way towards an answer [45]. The difference between experts and novices is easily framed in the context of constructivism; experts have a thorough understanding of the laws of physics, and thus feel comfortable starting from first principles. Novices often do not possess robust and accurate models of the laws of physics, so they need some sort of example by which they can relate the material to something they already know.

Larkin et al. described how expert problem solvers not only have a deep knowledge of their respective field, but also a sophisticated knowledge of the patterns that underlie seemingly different topics within the field. This allows an expert to identify important points in a problem and relate it to relevant parts of his or her knowledge base in a fraction of a second. When a person's knowledge of the field and its underlying patterns reaches a certain threshold, others might claim that he or she has an "intuition" about the field [45].

Chi, Glaser, and Rees conducted a study of how expert and novice physicists organize different problems. They found that expert physicists group seemingly different problems by their underlying fundamental principles (e.g. conservation of energy). On the other hand, novices grouped problems by surface features (e.g. the mention of an inclined plane in two problems) [46]. These differences in grouping problems are indicative of the different features that experts and novices identify in their initial analysis.

Another study by Chi, Feltovich, and Glaser found that experts represent physics knowledge in different ways. Experts conduct a qualitative analysis of a given problem before working with the appropriate equations. On the other hand, novices often start working with equations before reaching a solid understanding of the underlying principles governing the problem [47].

Expert problem solvers approach problems in a very different way from novices. They start with fundamental principles and work their way up to a solution. They do not focus on surface features in problems like similar diagrams or similar vocabulary. They have a deep knowledge about their field of study as well as an underlying understanding about the connections between seemingly different topics.

2.5 Scaffolding in Physics

Students need to contend with a variety of outside influences in the classroom. For example, they need to figure out what specifically is important to learn and how to make sense of new knowledge in the most efficient way possible. Additionally, STEM students need to be able to keep up with advances in their field and use contemporary knowledge to solve complex problems. Ultimately, professors would like their students to become independent learners who will continue to study on their own with limited support. Scaffolding provides an environment where students can systematically build an understanding of concepts, all the while gaining independence in developing their problem solving skills [48].

Within the context of education research, scaffolding refers to a variety of techniques that are used to systematically move a student towards a coherent understanding of the material. Teachers provide successive levels of temporary support that students can use to structure their learning. These levels of support are removed when the student no longer needs them. Some simple examples of scaffolding include [49, 50]:

- Using physical objects to simplify abstract concepts (e.g. teaching addition by counting pennies)
- Planning class work that utilizes hints and prompts from previous assignments
- Organizing group activities between students of different skill levels so that they might learn from each other

Physics instructors need to scaffold both the learning of qualitative concepts as well as problem solving techniques. Even though quantitative problems are more common than qualitative problems in introductory physics, cognitive science has shown that students need a solid background of qualitative knowledge in order to effectively work with equations [51]. Neto and Valente assert that a meta-cognitively oriented approach to problem solving promotes "a synergistic interaction between the scientific concepts and the thinking skills" [52].

Although much of the current research is in agreement with the idea of helping students build a string background of qualitative knowledge before engaging students in manipulating specific equations, it is worth noting that results are not always clear cut. Dukes et al. noticed that a conceptual understanding of physics material did not improve their students' abilities to solve a quantitative problem, but working through a quantitative problem improved scores on conceptual problems [53]. Nonetheless, CITA will provide a qualitative background before working with any equations.

A qualitative background of physics concepts is not the only method of scaffolding physics problem solving. Leonard et al. found that teaching problem solving strategies helped students identify major principles that could be applied to solve specific problems. It also helped students remember major physics principles even after the course had ended [54]. Systems like C3PO [55] use this same strategy of teaching students problem solving strategies in order to enforce physics concepts. This system will be explained in more detail later in this chapter.

The wording of a homework problem can lend itself to certain styles of solution. Heller et al. found that groups of students were more likely to use effective problem solving strategies when they were given context-rich problems to solve as opposed to standard textbook problems. The term context-rich refers to a challenging problem that requires a solid understanding of concepts and a logical problem solving strategy, rather than a trick or a single equation. These problems are rooted in real-world situations [56].

Mazur elaborated further by noting that standard end-of-the-chapter problems can be as ineffective as passive lectures in the teaching of physics concepts. Students often solve these traditional problems by identifying equivalent problems that they have solved before and applying the same techniques as before. This strategy typically does not promote learning [57].

Finally, I note that instructors can creatively use problems as a scaffolding unto themselves. Van Heuvelen et al. created a novel problem structure by literally turning physics questions backwards in a game called Physics Jeopardy. A physics problem in Physics Jeopardy introduces an equation or phrase that describes a physical process. The student must build a representation of the problem using a few short phrases, a picture, or followup equations. In other words, the student tried to represent a physical process in a variety of ways using diagrams, equations, and graphs [58].

Scaffolding in physics problem solving can be approached in a number of different ways. There is no "silver bullet" to scaffold a lesson. Scaffolding, at its core, is context dependent. Immersive CITA adopts a more traditional approach to scaffolding by first introducing qualitative concepts and then following them up with any appropriate equations.

2.6 Multimedia Learning

Earlier in this paper, I discussed the educational philosophy of constructivism and how it relates to physics education. Then, I discussed the differences between master and novice students and how scaffolding can help a student bridge the gap from novice to master. In this section, I dive into the details of how to actually structure an online tutorial system using Mayer's cognitive theory of multimedia learning as the basis for my decisions on how to structure each tutorial in the system [59].

2.6.1 Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning is rooted in the "multimedia principle" which states that people learn more from words and pictures than from words alone. Mayer suggests that each person has two channels which can take in and process information simultaneously - an auditory channel and a visual channel. Each channel has a finite capacity, similar to Sweller's cognitive load theory [44, 60]. By utilizing each channel to its fullest capacity, one can reinforce important concepts in a lesson.

It is important to note that the information that is taken in by each channel is not processed in isolation. True to the philosophy of constructivism, it is filtered, organized, combined, and integrated with one's prior knowledge to produce new models of the world. Thus, simply adding new words or pictures to a lesson will not necessarily achieve greater learning gains. The components of the lesson must work in harmony.

Mayer outlines twelve design principles that can be used to create effective multimedia tutorials. They are described below [61].

2.6.2 Mayer's Twelve Design Principles

Coherence Principle

The coherence principle states that extraneous words, pictures, and sounds only serve to distract the learner from the main point of a lesson. This is because each of information channels discussed above has a limited capacity; they can become overwhelmed when too many images or words appear at one time. Sanchez and Wiley take this a step further, noting that extra words and pictures are especially distracting to novices who are struggling to learn the concepts for the first time [62].

The CITA system truly takes the coherence principle to heart. As a student traverses an online tutorial, he or she will focus on a short paragraph (usually three or four sentences), a diagnostic question, and a simple diagram (if applicable). Students can scroll up the page to view previous steps in the tutorial, but their attention at any one given time is on a very focused part of the analysis.

Signaling Principle

The signaling principle states that people learn better when strategic cues highlight some of the essential concepts during the multimedia lesson. For example, a video may highlight the circumference of a circle as the narrator describes the relationship between circumference and diameter. This is an active area of physics education research; many groups are exploring how different types of cues are interpreted by students as they learn new concepts [63–65].

The CITA system makes extensive use of visual cues to guide students through problems. For example, bold or italic text highlights major vocabulary terms, colored boxes highlight correct or incorrect answers, and different shades of gray in the background indicate how deep into a tutorial a student has gone. However, it should be noted that these cues are used to simplify the traversal of a potentially complicated tutorial. They are not necessarily used to teach new concepts.

Redundancy Principle

The redundancy principle states that people learn more from graphics and narration than from a combination of graphics, narration, and text. There is a common belief that displaying text on a screen and narrating the words is better than doing one or the other. However, this is not necessarily true. The combination of text and narration is redundant; thus, it is best to eliminate any extraneous information in a multimedia lesson as per the coherence principle. The tutorials do not include any narration, so I do not have to worry about the redundancy principle here.

Spatial Contiguity Principle

The spatial contiguity principle states that people learn better when related words and pictures are presented close to each other rather than far from each other. Clark and Mayer provide a number of examples where the spatial contiguity principle might be violated in a multimedia lesson. Below are a few that pertain to online homework systems and online tutorials [66]:

- Separating text and graphics due to the need to scroll along a web page
- Separating diagnostic questions from their feedback
- Separating the main question with its corresponding tutorial (if information from the main lesson is used in the tutorial)
- Separating the navigation directions (i.e. click here, press this button) from the tutorial itself

The spatial contiguity principle was followed in the design of the CITA tutorials. Related graphics and content were always place close to each other to avoid excess scrolling - even if it meant that there was a slight amount of duplication of diagrams or text.

Temporal Contiguity Principle

The temporal contiguity principle is very similar to the spatial contiguity principle, except for the fact that it describes time-separated content. It states that people learn better when related graphics and text are presented simultaneously rather than in succession. This is because the learner needs to simultaneously use their working memory to store past information and process new information.

The structure of the tutorials adheres to the temporal contiguity principle. All of the content that is required for a given step of the analysis is presented in the current step of the tutorial. Additionally, if students wish, they can easily recover previous information about the problem by scrolling upwards to previous steps on the page.

Segmenting Principle

The segmenting principle states that people learn better from a multimedia lesson that is broken into short segments rather than a single monolithic unit. The pace at which these short segments are presented must be determined by the student. The reasoning behind this principle goes back to the dual-channel model of multimedia learning. If a student has a chance to control the pace of a multimedia lesson, he or she can take short breaks to digest each segment and avoid cognitive overload. However, if information is continuously being fed to a student (like in the case of a video lecture without a pause button), he or she cannot stop to think about what has just been said, and cognitive overload can occur.

I tried to make sure that students can traverse CITA at their own pace. Each step in a tutorial consists of a short passage of text, a diagnostic question, and a diagram if applicable. Students can read the text at their own pace, referring to external resources without any time pressure. The diagnostic questions divide each step of the tutorial, forcing the student to slow down and make sure that he or she fully understands the explanation before moving onward.

Pre-Training Principle

The pre-training principle states that people learn better from a multimedia lesson when they know the accompanying vocabulary and concepts beforehand. The reason for this is that the student does not have to spend energy identifying unfamiliar terms while simultaneously learning how they fit together. Instead, the student can focus all of his or her effort on the main content of the lesson.

The pre-training for the tutorials comes from the class lectures and assigned readings. Students must attend or watch lectures twice per week to learn about the major concepts in electromagnetism. This means I could use general physics jargon in the tutorials. For example, a student may first learn about electric charge, electric field, and Gauss's law from their lectures and readings. Then, they will learn about the subtleties of how symmetry affects Gauss's law in the tutorials.

Modality Principle

The modality principle states that people learn more from an animation paired with a narration rather than an animation paired with on-screen text. In other words, spoken word is more effective than written word in an animated multimedia lesson. This is because the learner might not have enough time to identify and process written words on the screen in addition to the animation.

On first glance, it might seem like I am violating the modality principle in the design of the tutorials. After all, I exclusively use text rather than narration. However, it is important to note that the modality principle applies to animations, of which the tutorials have none. Additionally, recent research has found that the modality principle might only apply in certain situations [67]. Clark and Mayer elaborate on some of the situations where the modality principle does not apply in general. All of these specific situations apply to the system and student body [66]

- The text is long and complex.
- The text contains technical terms or symbols.
- The text is not in the learners native language (foreign students only).
- The material is paced by the learner.

For these reasons, I can safely ignore the modality principle in the development of CITA.

Multimedia Principle

The multimedia principle states that people generally learn better from a combination of words and pictures than from words alone. This is because pictures are processed through the visual channel while words are processed through the auditory channel (you read the words in your head and "listen" to the internal dialogue). Thus, it is preferable to have a combination of words and pictures in a multimedia lesson rather than just one or the other.

Pairing diagrams with explanations is a staple in physics, and the tutorials are no exception. Almost every tutorial either starts off with an explicitly drawn diagram or makes the student draw their own diagram which will be used in the analysis. Furthermore, pictures and plots are used extensively throughout all of the tutorials to explain abstract points.

Personalization Principle

The personalization principle states that people learn better when the multimedia lessons are in a conversational style rather than a formal style. A spoken or written tutorial should address the audience politely and personally rather than convey everything in a "sterile" way. This means that it is often appropriate to use the first or second person within a multimedia lesson, even though this is very different from the standard third person narrative that is the style of traditional textbooks.

Clark and Mayer greatly expand on this idea in their research. The use of a conversational style of writing or speaking more closely resembles human-to-human conversations. Although interacting with a computer is not the same as interacting with a human, an informal or conversational has repeatedly shown to be more effective than a formal third person style of writing [66].

All of the written tutorials in the CITA system are written in an informal, conversational style. The author often addresses the reader directly (i.e. what would you do at this point?) and includes himself in the conversation (i.e. many students in my class often do this...). The only place where the text is written in the third person formal style is the question itself since this is how it would appear on exams.

Voice Principle

The voice principle states that people learn better when the narrator in a multimedia lesson has a friendly human voice rather than a machine-like voice (for an example of a machine-like voice, you can refer to the dictation program paired with most operating systems [68–70]). The voice principle can also be extended to foreign accents - students tend to learn better when the narrator has a familiar accent rather than a foreign accent. This principle does not apply directly to the system since I do not have any narration in the tutorials.

Image Principle

The image principle states that people do not necessarily learn more from a multimedia lesson when the speaker's image is paired with the dialogue. This could be interpreted as an extension of the coherence principle - the image of the narrator is extraneous information that the student has to contend with in addition to the lesson itself. Like the voice principle, the image principle does not apply directly to the system since I do not have any narration in the tutorials.

2.7 Existing Online Homework Systems

Numerous studies have shown that online homework is just as effective or more effective than paper-and-pencil assignments [71–73]. There are a number of online systems that perform various tasks in the classroom. Some are dedicated homework systems that test student performance and provide feedback. Others pair diagnostic questions with lessons in an attempt to teach as well as test content knowledge. Course management systems (CMSs) provide a complete tool for educators by pairing a homework system with classroom management tools (i.e. grade books and messaging abilities).

The systems are further parameterized by supplemental features. What subject(s) does the system focus on? Does the system include resources for an external concept inventory (CI)? How much will the system cost students and schools? Are there additional resources to create tests and quizzes based on the homework questions? The list goes on.

In this final section, I describe some of the other online homework systems that are out in the market today. All of the systems present questions to students and record their answers. However, many of them differ in subtle ways. This is certainly not an exhaustive list - it merely touches on systems that are either widely in use or are pursuing interesting paths of research. I describe systems that have unique ways of approaching online learning.

2.7.1 WebAssign

WebAssign is a well-known name in the field of introductory science education. According to their website [74]:

> WebAssign is a powerful online instructional system designed by educators to enrich the teaching and learning experience. WebAssign provides extensive content, instant assessment, and superior support.

This online homework system is well-established and well-tested with a variety of resources for instructors that include: questions tied to a variety of physics textbooks, a sophisticated grade book with some filtering capabilities, a small selection of validated concept inventories, and integration of videos into problems (e.g. students have to calculate the kinetic energy of a roller coaster from a video). WebAssign's enormous library of questions contains a variety of types including short answer, multiple choice, numerical, and check box.

2.7.2 McGraw Hill Digital Platforms

McGraw Hill released a series of software products called Digital Platforms that can be used in a variety of subjects including mathematics, science, and foreign languages [75]. These products provide everything from lecture recording services to personalized electronic grade books to online homework systems. The two systems most relevant to online homework as I discuss here are McGraw Hill Connect and McGraw Hill LearnSmart.

McGraw Hill Connect is an online homework system that offers many of the same features as WebAssign. A teacher using Connect can create online homework assignments from a large database of questions and administer them to the class. Grading and analytics are done in the background, so the instructor does not have to spend as much time reviewing student's assignments. McGraw Hill Connect references media-based e-books when students need help with topics on the homework.

McGraw Hill LearnSmart is an adaptive learning product within McGraw Hill Connect that collects a wealth of data about students as they answer diagnostic problems. For example, it tracks their correct responses, the reported confidence in an answer, and the time spent on a given question. Then, this information is fed into a proprietary algorithm that adjusts the instruction that a student receives.

2.7.3 Sapling Learning

Sapling Learning is an online homework system that shares many of the same qualities as WebAssign and McGraw Hill Connect. It too has a sophisticated grade book to track student progress and a large number of physics problems that instructors can use in their classes. However, it does not have the huge amount of third-party textbook support that WebAssign currently has [76].

What Sapling Learning lacks in third party support, it makes up for in personal service. The company has stated that they are committed to providing educators with “Tech TA” that can help the instructor plan his or her course [77]:

> We match educators with a Sapling Learning Technology TA a Ph.D. or masters-level subject expert who provides collaboration, software expertise, and consulting to tailor each course to fit your instructional goals and student needs.

The Sapling Learning testimonial page indicates that this feature is highly rated amongst their clients and sets them apart from other companies.

2.7.4 Mastering Physics

Mastering Physics is another online homework system that is similar to WebAssign and Sapling Learning in its base features. What sets Mastering Physics apart from other homework systems is the adaptive content [78]. Mastering Physics claims that since every student learns at a different rate, their adaptive learning exercises assess student activity in real-time in order to provide personalized feedback to students. These Dynamic Study Modules ask students to answer a set of questions as well as rate how confident they are with their answer. The module repeats itself with variations of the questions until the student can answer all of them accurately and correctly. The company recommends that these modules be used both before and after a homework assignment in order to solidify knowledge of the concepts.

2.7.5 The Expert TA

The Expert TA is an online homework and tutorial system that is not affiliated with any specific textbook [79]. It contains all of the usual features that one might expect from an online homework system - a sophisticated assignment management GUI, detailed grade reports, adjustable grading scales, and a large library of physics problems that can be incorporated into homework assignments with little effort. However, The Expert TA has set itself apart from other online homework systems in two big ways.

First, the company has spent a lot of time and energy developing their library of symbolic questions. Most physics homework systems include some symbolic questions (including CHIP); they are usually quite restrictive or clunky. A student can choose from a list of symbolic answers or enter an equation with a very specific format (e.g. the system might interpret $1/x$ and x^{-1} as two different answers). However, The Expert TA has designed an Equation Entry Palette that dynamically responds to student input and makes entering symbolic expressions quite simple. Additionally, they have built a proprietary mathematics engine that can recognize the similarity of different symbolic expressions.

Finally, the company has a well-developed analytics back-end that collects details about how students and instructors use the system. The Expert TA can use this information to design focused feedback on often missed questions as well as suggest potential questions for instructors to use in their classes.

2.7.6 LON-CAPA

LON-CAPA is an acronym that stands for the Learning Online Network with a Computer Assisted Personalized Approach. It is a free, open-source, homework system paired with a CMS. Unlike many online systems, the homework system and CMS were actually born out of two different projects - LON and CAPA [80]

CAPA is meant to provide students and instructors with homework sets as well as problems for exams. However, these are not just bare-bones questions. Students are given instant feedback right through the CAPA GUI that can be used to adjust their answers. The Lecture Online project was originally intended to serve physics education content to instructors through the internet. The two systems combined in 1999 to create LON-CAPA (note how Lecture Online became Learning Online). They were some of the first systems of their kind.

The biggest strength of LON-CAPA is its database of open content. When instructors create new materials, they have the option to share that content with the rest of the LON-CAPA network. Their work gets pooled into a shared cross-institutional resource library. Later, when a new instructor wishes to access some open content, he or she can pick and choose whatever is appropriate for the course. In other words, LON-CAPA allows instructors to share and retrieve content at various levels of detail in order to build a custom homework experience for their students.

2.7.7 FlipIt Physics

FlipIt Physics (previously called Smart Physics) is a "complete course solution" for calculus and algebra based physics classes [81]. Rather than focusing simply on homework alone, FlipIt Physics attempts to combine pre-lecture, lecture, and homework into one seamless learning experience. In fact, the name "FlipIt" is a reference to the "flipped classroom" approach used by this system.

First, students watch a narrated, multimedia pre-lecture that introduces some of the core physics concepts that will be presented in an upcoming lecture. The pre-lecture contains a small set of questions that ensures students are following the narration; it finishes with a formative Bridge Question that summarizes the main ideas of the material and connects the pre-lecture to the upcoming lecture. Instructors can use the results of the Bridge Question to pinpoint student misconceptions of the material and fine-tune the upcoming lecture. This process is similar to Just in Time Teaching except that the pre-lecture assessment is more directed by the instructor [82].

Afterwards, students practice what they learned in the pre-lecture and lecture within the homework. FlipIt Physics uses what are called "Interactive Examples" as scaffolding in order to teach students physics content and good problem solving skills.

2.7.8 Minds on Physics Learning Modules

Minds On Physics is a collection of 15 learning modules that are structured to teach students introductory physics topics. The modules consist of 135 assignments containing over 1300 questions, each designed to encourage reflection and review.

Recent literature has suggested that reflection of an answer is extremely important in the learning process [83]. The authors of Minds on Physics have designed the problems such that superficial answers are “quickly challenged” [84].

Unlike many other homework systems, Minds on Physics does not include any built-in grade book or score tracking system (as of the time of this writing). Rather, instructors can use a combination of an instructor code and a student code to retrieve information about the completion of the assignment. Then, the instructor can enter this information into an external grade book.

2.7.9 Customizable Computer Coaches for Physics Online

While most online homework systems focus on the physics content of the homework problems, the physics education researchers at the University of Minnesota have gone in a different direction. Although they would certainly agree that physics content is important, they stress that the process of thinking through and solving a complicated problem is more valuable than having a superficial knowledge of concepts. Thus, their physics education research group has developed a set of Customizable Computer Coaches for Physics Online (C3PO) [55].

As of now, these coaches have only been developed for a handful of physics problems. They feature a split-screen GUI that displays the problem text, a diagram when appropriate, the current step of the analysis, and an organized list of previous steps. Students move through an organized problem-solving framework that is a slight variation of what was put forth by Polya:

- Focus the Problem
- Describe the Physics
- Plan the Solution
- Execute the Plan
- Evaluate the Answers

2.7.10 Learning Management Systems

A learning management system is a software suite that can be used to organize and run a classroom. Generally a learning management system is focused on the administration of a class rather than the teaching of new materials. However, systems like Moodle and Blackboard have created resources for instructors to create small-scale assignments directly through the LMS. These assignments do not offer the sophisticated teaching resources that are discussed above, so they are generally useful for small diagnostic assignments like pre-class quizzes or reading comprehension assessments.

Blackboard Learn is a management system developed by Blackboard Inc. It excels in two main areas: (1) adding an online element to courses that were previously face-to-face and (2) creating and administrating an online-only course with no face-to-face meetings. As of now, it is one of the most popular systems available due to an open architecture, integration with other homework systems, and scalable design. Blackboard can be installed and maintained on a school's local servers, or it may be hosted by Blackboard ASP Solutions [85].

Moodle (Modular Object-Oriented Dynamic Learning Environment) is a free, open-source learning management system with over 68 million users world-wide [86]. It was designed to help educators create their own online courses with a focus on interaction and collaboration. Moodle uses a simple drag-and-drop interface along with specialized text editors to allow educators to build their own web pages. Moodle classrooms are based on a social constructivist learning philosophy, and thus, provide the educator tools to build social learning assignments like wikis, surveys, and chat rooms. Schools can download and host the Moodle software on their own campus, or they can hire a Moodle partner (e.g. Moodle Rooms [87]).

2.7.11 Massive Open Online Courses

This final category describes not one system, but a collection of online systems that teach physics and administer homework. Massive Open Online Courses (MOOCs) have grown in popularity over the past few years. They are usually characterized by large class sizes and unlimited participation due to the fact that they are hosted entirely (or almost entirely) through the Internet [88]. Although MOOCs are first and foremost geared towards teaching content, they usually are paired with an online homework/assessment system to validate that learning has actually occured.

Some of the most popular companies that provide MOOCs to students include Khan Academy [89], Udacity [90], edX [91], Coursera [92], and Academic Earth [93]. Most of these companies are paired with one or more universities so that professors can create new content. Oftentimes, the professors who want to use this content for their own, small-scale class might restructure a MOOC as a SPOC - a Small Private Online Course.

2.8 Future Directions

Sir Tim Berners-Lee proposed the idea for the World Wide Web on March 12, 1989. Google was incorporated on September 4, 1998. Facebook was founded on February 4, 2004. As hard as it might be to believe, the web as we know it is still a teenager. Thus, researchers are exploring a vast and unknown area of education.

One of the current directions in online education research is the merging of "big data" with traditional assessment techniques. Researchers are leveraging student demographic information along with real-time performance on a homework problem to provide focused, relevant feedback to the student. As of now, little is known about what kinds of tutorials aid and detract from learning.

Another direction of research is the merging of social networks with traditional tutorials. Many researchers are building social learning platforms in the hopes that the social constructivist learning philosophy will lead to large gains in student learning.

For example, students might be broken up into small groups and tasked with reading a textbook passage as a team. Questions and comments made by one student can be viewed and expanded upon by many others. Although much research has been done on learning in groups, we have never seen group learning on such a massive scale.

There are many other open questions in this field. We have barely scratched the surface on what is truly possible in the realm of online education.

3. METHODS

3.1 Preface

Chapters 4 and 5 along with Appendix A each contain a methods section since they are papers that are submitted for publication. However, I have compiled and summarized the details of the methods of this project in this chapter to parallel a traditional dissertation format.

First, I give an overview of the research design and the participants involved. Second, I describe the structure of the CITA system and its schedule of development. Third, I elaborate on the qualitative and quantitative methods that I used to analyze the data. Finally, I end with connections between the methods, collection procedures, and the role of the researchers.

3.2 Research Design

This research project used a mixed methods design. Mixed methods research involves collecting, analyzing, and integrating both qualitative and quantitative data in order to gain a better understanding of a research domain than either approach alone. The mixing of qualitative and quantitative data can be very powerful for a number of reasons. First, the synthesis of data from different sources leads to greater confidence in the validity of the conclusions. Second, the answers to the research questions merge a number of perspectives, leading to a more complete conclusion with fewer "gaps" in the analysis. Finally, any pre-existing assumptions from the researchers are less likely to influence the results due to the fact that the conclusions must support both qualitative and quantitative data [94].

My project benefited from a mixed method design because it allowed the research team to explore the subtleties of how the CITA tutorials influenced student performance. Scores on exams and quizzes are certainly useful for determining learning gains, but they do not yield any information about how students used CITA. The qualitative aspect of the mixed methods design yielded insight into motivations for completing homework assignments, problem solving strategies, and the utility of the CITA tutorials compared with other external resources (e.g. websites, the textbook, class notes, etc.).

This research project took place between the spring semester of 2015 and the summer semester of 2016 at Purdue University. It focused on the second semester, introductory physics courses for students pursuing an engineering degree - PHYS 24100 and PHYS 24100D. The "D" signifies a distance learning (i.e. online) section of the course. As described in Chapter 1, the students in each of these sections covered the same content, but it was presented in either an on-campus or online setting.

3.3 Research Participants

Students had the option of following a CITA tutorial in order to learn how to solve a given homework problem. Alternatively, they could skip over a tutorial paired with a homework problem completely. Whether a student decided to use the CITA tutorials or not, he or she was asked to complete a demographics survey and exit survey that probed their usage of the system. At the end of the semester, students were given the option to elaborate more on their opinions of the system through voluntary focus group interviews.

Thus, students were only required to complete the surveys during the semester; CITA tutorial use and focus group sessions were completely voluntary.

3.3.1 Previous Skills

Without an understanding of the previous skill levels of the PHYS 24100 students, I could not make accurate predictions about how the CITA system affected their performance. Thus, I conducted a study of student's previous grades in physics and calculus in order to determine if the online and on-campus sections of PHYS 24100 contained significantly different populations of students. Students were asked to complete a demographics survey at the beginning of the semester that asked about their prior performance in PHYS 17200 and calculus. Students had the option of answering "I Prefer Not to Disclose" if they wished to keep their information private. However, even with the option not to report any information, most students provided answers to all of the questions. The information was categorized by semester, section, and answer to the question. The populations were compared using chi-squared tests of significance.

The fall semester is considered the on-semester while spring and summer are considered the off semesters; this means that students who are on schedule with their plan of study would take PHYS 24100 during the fall. Thus, my hypothesis was that the students in the fall semester would report higher physics and calculus grades than those in the spring or summer semesters. Additionally, I hypothesized that the online sections of PHYS 24100 would be filled by the students who waited until the last minute to sign up for classes, and thus, are generally not as successful as their more motivated colleagues. In other words, self-reported grades in the on-campus sections would be higher than self-reported grades in the online sections.

My first hypothesis from above was partially correct. PHYS 17200 grades are generally higher in the fall semester than in the spring or summer. Chi-square tests show a statistically significant difference between the reported grades. However, calculus grades do not show any statistically significant difference between any of the semesters. My second hypothesis was incorrect. Both the self-reported PHYS 17200 grades and self-reported calculus grades were not significantly different between the online and on-campus sections of PHYS 24100 within any given semester.

Further inspection of the data showed that about 22% of students in the summer of 2015 were retaking the class because they received a low grade (D or F) in a previous semester. The spring and fall semesters were much different - about 11% of the students in the spring of 2015 and 5% of the students in the fall of 2015 were retaking the class due to a low grade.

I also found that the distance learning sections of PHYS 241 attracted a disproportionately large number of female students. Online sections were about 60% male and 40% female during the fall 2015 and spring 2015 semesters. However, they were about 55% male and 45% female during the summer of 2015. This is much different from the on-campus enrollment, which ranged between 75-80% male and 20-25% female during the spring 2015 and fall 2015 semesters.

Students reported similar ranges of calculus grades within every section and semester. Additionally, students reported similar range of PHYS 17200 grades in most cases - the one exception being that PHYS 17200 grades were higher in the fall 2015 semester when compared with other semesters. I found that the spring and summer semesters of PHYS 24100 attracted more students who were retaking the class, and the on-campus and online sections of PHYS 24100 attracted more female students than male students.

3.4 Design of the CITA System

3.4.1 Structure of the Tutorials

CITA is built on top of the CHIP homework system. CHIP is based on an online homework system called CPlite which was originally developed by the faculty at the University of Illinois in Urbana-Champaign. It was adopted by Purdue University and has been under constant extension and revision [13]. The research group decided to expand the CHIP system rather than build a new homework system from scratch for a variety of reasons (see Chapter 1).

CITA provides a series of successively more detailed tutorials that target students with a variety of skill levels. The system is divided up into three parts: Shallow CITA, Immersive CITA, and Postscripts. These parts are not meant to work independently of each other. Rather, they work together to take a student through Polya's problem solving process in a controlled way [39].

An example of Shallow CITA is shown in Figure 3.1. A student is given a homework problem that counts for a grade. If the student makes an easily identifiable mistake on the problem a red box will pop up and offer specific feedback on how he or she should correct the mistake to proceed. Shallow CITA assumes that a student generally understands the major concepts in the problem and is close to a correct solution. If the system cannot identify the mistake made by the student, it suggests that he or she move onto an Immersive CITA tutorial.

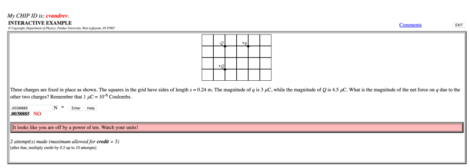

Figure 3.1. A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error.

Immersive CITA was built to provide detailed tutorials about the specific concepts within a given homework problem (see Figure 3.2). In the fall of 2015, we paired step-by-step tutorials with 109 of the problems on the homework - just under 80% of the total problem count. Problems that consisted of more than three graded parts were not given an Immersive CITA tutorial since the multiple parts already hinted at a path to the solution. Secondly, true or false questions were not given a tutorial due to the fact that the question had only two potential answers. Finally, problems within a homework assignment that repeated similar concepts were given truncated tutorials that outlined the solution to the problem in a few sentences or referred the student to a relevant Immersive CITA tutorial in a different problem.

During the spring semester of 2016, the CITA development team (henceforth "we") updated most of the truncated tutorials mentioned above to detailed step-by-step tutorials. This meant that there was repetition of concepts within the tutorials within a single homework assignment. However, no tutorials were repeated verbatim. Additionally, we added a filter to 74 of the Immersive CITA tutorials (slightly over 50%) that asked students to rate their confidence with the problem and decide if they wanted a detailed, step-by-step tutorial or a faster, general overview of the solution. The system would then filter the student body into one path or the other. Wherever possible, we tried to keep the detailed step-by-step tutorials the same between the fall 2015 and spring 2016 semesters so that we could make consistent comparisons. Again, about 80% of the homework problems included some form of tutorial in the spring of 2016.

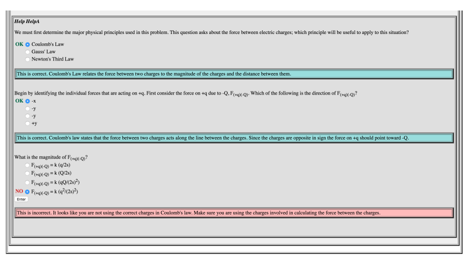

Figure 3.2. A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion.

We wanted students to work through the solution of a problem rather than just read a transcript outlining the solution. Thus, each step of an Immersive CITA tutorial consists of just a few sentences, followed by a question. In this way, the solution is methodically worked out by the student. Each Immersive CITA tutorial followed a similar structure based on Polya's problem solving strategy. First, we spent a few steps describing some of the major conceptual ideas contained in the solution. Next, the tutorial devised and worked through the solution, referring to lecture slides, class notes, and previous steps of the analysis.

Figure 3.3 shows the final part of CITA: the Postscript. After a student correctly answers the homework question, the Poscript reviews Polya's final step in the problem solving process. We ask students to think about the work that they just completed by outlining variations of the problem. For example, if a student determined the electric field at a certain point in space was equal to zero, a Postscript might ask what would happen if the charge distribution was perturbed from its current location. During the spring of 2016, each of the postscripts included a "Real World Application" to outline where the concepts in the problem could be used in the realm of engineering.

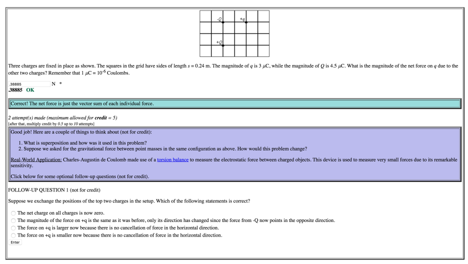

Figure 3.3. A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer.

3.4.2 Schedule of Development

Development of the CITA system began in the spring of 2015. Over the next year and a half, we used a design-based research method to build, assess, and fine-tune the tutorials. The schedule of development is shown in Table 3.1.

Version 1 of the system was our first foray into development. We proceeded slowly, building the structure that would be used in the branching system and ensuring that the tutorials worked for students. We were able to provide CITA tutorials for approximately 50% of the homework problems (while the remaining problems contained no help of any kind). The original "interactive examples" that were developed by UIUC were given Shallow CITA feedback and are included in the statistic.

Version 2 of the system contained the step-by-step tutorials as described earlier. This version of the system was designed to serve two purposes. First, it was an intermediate step between the initial and final system so that we could break the development into two roughly equal parts. Second, this version allowed us to run an analysis of a more traditional tutorial system in order to get some baseline data. During the fall semester of 2015, there was a problem in development, so the online courses were given Version 1 of the system while the on-campus classes received Version 2. This anomaly has been taken into account in the analysis.

Version 3 of the system consisted of problems that offer the choice between two different levels of scaffolding (as described in the previous sections). This is the final goal of this particular research project.

3.5 Quantitative Procedures

3.5.1 Statistical Methods

The quantitative analysis made use of linear models, Student's T-tests, ANOVA analysis, and chi-squared tests where appropriate. I used the traditional cutoff of $p = 0.05$ for tests of statistical significance along with Cohen's guidelines for interpreting effect sizes [95]. In most cases where ANOVA and t-tests were used, the data did satisfy the necessary assumptions:

1. Independence of Cases
2. Normality of the Data
3. Equality of Variances

In any cases where at least one of the three conditions was not satisfied, I used non-parametric tests to verify that any statistically significant differences in the means

Table 3.1
The schedule of development of the CITA system. In the spring semester of 2015 and prior, the CHIP homework system included 29 "interactive examples" that were originally developed by the faculty at UIUC (out of 139 homework problems total). These 29 problems were paired with step-by-step tutorials and made up approximately 21% of the homework problems seen by students.

Semester	**Section**	**Version**
Spring 2014	Campus	None
	Online	None
Summer 2014	Online	None
Fall 2014	Campus	None
	Online	None
Spring 2015	Campus	None
	Online	None
Summer 2015	Online	1
Fall 2015	Campus	2
	Online	1
Spring 2016	Campus	3
	Online	3
Summer 2016	Online	3

was actually true. The most common non-parametric tests that I used were the Kruskal-Wallis and Welch tests [96].

Measures of effect size included Cramer's V, eta squared, and R^2 where appropriate [97, 98]. Cramer's V is a way of calculating the effect size of contingency tables from a chi-squared calculation. First, one calculates the chi-squared statistic of the contingency table (χ^2). Then, Cramer's V is equal to:

$$V = \sqrt{\frac{\chi^2}{NM}} \tag{3.1}$$

where N is the sample size and M is the smaller of the number of rows in the table minus one or the number of columns in the table minus one.

Eta-squared is defined as the proportion of the variance accounted for by each of the interactions in an ANOVA study. It is defined as [99]:

$$\eta^2 = \frac{SS_{effect}}{SS_{total}} \tag{3.2}$$

where SS stands for the sum of squared errors of the effect of interest (numerator) and all effects, interactions, and errors (denominator).

Finally, R^2 is the goodness-of-fit in a linear regression of two variables. It is a unitless fraction between 0.0 and 1.0. A value of 0.0 means that knowing one variable does not help you predict the other (i.e. there is no linear relationship between the variables). A value of 1.0 means that all points lie on a straight line (i.e. knowing one variable lets you predict other perfectly).

3.5.2 Exam Scores

Homework scores predictably increased between spring 2015 (pre-CITA) and spring 2016 (post-CITA). However, this increase does not tell us how students perform when the scaffolding is removed. One of the ways that I assessed student learning gains was through their exams. A few of the exam questions during the fall of 2015 and spring of 2016 appeared on exams in previous semesters. Thus, I used these questions to assess differences between students in different semesters. I also compared students within a given semester to see how those who extensively used CITA compared to those who did not.

3.5.3 Brief Electricity and Magnetism Assessment

The BEMA was developed by Ruth Chabay, Bruce Sherwood, and Fred Reif in 1997 [100]. Although it was originally designed to measure student retention of electricity and magnetism concepts three months to five semesters after completing an introductory electricity and magnetism course, it is now often used to analyze student learning between the beginning and end of the semester. The BEMA is a useful tool to assess the understanding of electricity and magnetism concepts that are covered in a college-level calculus-based introductory physics course.

The BEMA is a multiple choice test consisting of qualitative questions and a few simple calculations. Lin Ding et al. performed an analysis of the BEMA, showing that it is a reliable assessment tool for introductory electricity and magnetism courses [101, 102]. Although the CITA system does not focus on teaching a conceptual understanding of electromagnetism topics, we still used the BEMA to assess if there were any changes in this area.

3.5.4 Multi-Step Problem

Chabay and Sherwood demonstrated a technique to analyze how well students can solve a non-trivial, multi-step problem [103,104]. They gave their students a complex problem and tracked how far students made it into the analysis before getting stuck. Then, they plotted the curves for how many students made it to certain points in the problem before getting stuck.

We administered a non-trivial problem to on-campus students during their recitations and graded their responses with a standardized rubric (see Appendix 4.12). Students were first given a 20 minute review of the major topics in the problem by their teaching assistant. Then, the problem was given under test-like conditions (i.e. no books, no notes, no collaboration with others). Once again, this allowed us to assess how students fared without the scaffolding provided by CITA. The control group for this analysis was the spring 2015 on-campus students while the experimental groups were the fall 2015 and spring 2016 on-campus students.

3.5.5 Student Exit Survey

The PHYS 24100 faculty asked students to fill out an online survey at the end of the semester. It asked about their opinions of the CITA system as well as how they used it. This survey was required for a recitation quiz grade (as mandated by the internal review board at Purdue University). However, students had the option to select “I Prefer Not to Disclose” for each question.

3.5.6 Recording Clicks

As a student worked through the graded problem or an Immersive CITA tutorial, his or her progress was automatically recorded by the system. I reviewed the click history of the students in order to see how they traversed the different structures of tutorials.

3.6 Qualitative Procedures

3.6.1 Strategy of Inquiry

The qualitative side of this study analyzed student opinions about the CITA program as well as their general study habits. Qualitative data came from three main sources: (1) the final question of the exit survey, (2) student comments on the Piazza help system, and (3) focus group sessions at the end of the semester. After the second exam (when we started on the AC circuits unit) emailed the students to ask for volunteers for a one hour focus group session. The sessions took place at the student’s convenience in the two weeks before their final exam. After each session, we rewarded participants with a ten dollar gift card.

The strategy for analyzing the transcripts from the three sources consisted of a three-cycle coding plan [105]. In the first cycle, each of the transcripts was analyzed separately using attribute coding and descriptive coding. The purpose of the first cycle was to identify common responses given by the students to each of the questions asked in the sessions. The second cycle also treated the transcripts separately; it consisted of rounds of elaborative coding and pattern coding. The purpose of this cycle was to group the responses from the first cycle and identify overarching themes within a specific source. The final cycle of coding pooled all of the sources together. It consisted of evaluation coding and longitudinal coding to identify how comments compared between semesters.

3.6.2 Focus Group Sessions

Focus group sessions were ideal for this study. They are flexible since they can be used for exploratory, explanatory, and evaluative research. Additionally, they create a large volume of data with a range of viewpoints from all of the participants [106]. In this study, I conducted 11 focus group sessions with four scheduled participants in each group (six in the fall 2015 semester and five in the spring 2016 semester). Each one lasted one hour and was audio recorded. The audio recordings were transcribed for later analysis.

I used purposeful sampling to assign students to different groups. Patton defines this strategy as one where the researcher chooses specific participants for the study in order to gain insight about a specific phenomenon. Cases are selected because they are "information rich" [107]. When students signed up to volunteer for a focus group session, I asked them what their overall impression of the homework system was - "Generally Positive", "Generally Negative", or "Neutral". I then divided the participants into groups of four based on their reported overall impression of the system and whether they were in an online or on-campus section. I wanted to match similar opinions beforehand so that students were comfortable sharing both positive and negative opinions.

3.6.3 Analysis Procedures

Once I compiled the transcripts from the focus group sessions, Piazza comments, and written survey answers, I coded the data to find any underlying patterns. I used inductive analysis and creative synthesis as the analysis and reporting strategy. This pair of techniques entails starting from the specific details of an inquiry and slowly uncovering patterns in the data until a broad theory can be synthesized [107]. It allowed me to mesh together the opinions of the students in a controlled way.

My qualitative analysis used a three cycle system. First, I coded the transcripts in order to find specific words and phrases that highlighted student opinions. Next, I compiled these keywords into categories that described student opinions within a semester. Finally, I compared student opinions between the semesters.

3.6.4 Roles of the Researchers

A qualitative researcher must acknowledge that his presence in the experiment can alter the answers given by the participants. Thus, he must take time to determine how this might skew the data as well as how to minimize this effect [108]. In this study the principle investigators were Mr. Cyrus Vandrevala, Dr. Lynn Bryan, Dr. Andrew Hirsch, Dr. Hisao Nakanishi, and Dr. Laura Pyrak-Nolte.

Mr. Cyrus Vandrevala has been a teaching assistant for PHYS 24100 and PHYS 24100D since the fall semester of 2011. He has coordinated the course during summer sessions and helped develop the online sections, which included editing course videos and setting up the online Piazza classroom [109]. In order to prevent a conflict of interest, Cyrus has not taken a teaching assistant position in PHYS 24100 since the spring of 2015. Instead, he had a part-time assignment answering questions on

Piazza. This means that students were not be interviewed by any faculty member who assigned them a grade.

Dr. Hisao Nakanishi is the “father” of CHIP because he is one of its primary developers. Hisao Nakanishi developed the underlying structure upon which the branching tutorials were built. Although he was not directly affiliated with Electricity and Optics, he addressed student questions sent through the CHIP help system that pertain to administration issues. Additionally, he helped review all of the content that was uploaded to the website.

Dr. Laura Pyrak-Nolte has taught PHYS 24100 and PHYS 24100D since 2005 and coordinated the course since 2011. She is responsible for creating the curriculum, writing the exams, assigning final grades, and approving all changes to the homework system. Additionally, she was the professor that was recorded for the lecture videos shown to the online sections. Since Laura Pyrak-Nolte was the coordinator for PHYS 24100, she did not conduct any of the focus group sessions or interviews with the students.

Drs. Lynn Bryan and Andrew Hirsch provided the educational theory that supported this study. Neither of them were directly involved with PHYS 24100 or PHYS 24100D. It should be noted that Andrew Hirsch is one of the main instructors of PHYS 17200 - the introductory physics class that is a prerequisite for PHYS 24100 and PHYS 24100D. I used many of the techniques that he uses in the analysis of PHYS 17200 to assess PHYS 24100 students.

3.7 Quantitative and Qualitative Connections

The procedures outlined above were combined so that each set of results helped shape the next assessment. The research team employed a sequential exploratory strategy for the development and testing of CITA [110]. We conducted a phase of qualitative data analysis followed by a phase of quantitative data analysis. The two types of analysis built on each other and wove a detailed narrative.

Qualitative data was collected at the end of a semester through Piazza comments, an exit survey, and focus group transcripts. Then, I performed a small-scale qualitative analysis that guided the next step of CITA development as well as refine the quantitative measures. Next, the team collected new grade data while analyzing the grade data from the previous semester. These quantitative results were then used to adjust the qualitative procedures for the next iteration. This procedure was repeated four times over the course of the study.

3.8 Data Collection Procedures

CHIP leverages Purdue's security system to protect student records. Every student and instructor needs to sign into the system using a unique Purdue user name and password. The support staff has created a series of roles to determine the permissions for CHIP users (student, instructor, or administrator). Thus, there are electronic barriers preventing students and faculty from accessing data that they do not have a right to view.

The research group took additional measures to ensure that all of the data from focus group sessions was protected. We used an external company called Rev.com [111] to transcribe all of the interviews. The company only received audio data from the sessions with no sensitive or identifying information included in the recording. All of the recordings are stored on the Purdue University Research Repository system, which is protected by the Purdue's web security.

Finally, the research team only analyzed grade data after a semester had ended. This was advantageous to the project because the team had a complete data set to analyze (some students waited until the end of the semester to complete assignments). This also minimized a potential conflict of interest between the goals of the researchers and the goals of the students.

4. OVERVIEW OF THE CITA SYSTEM

4.1 Preface

Included in the paper below are an overview of the CITA system and the results of a study that examined: (a) who uses CITA; (b) how does the performance from Electricity and Optics students who use CITA compare to those who do not use CITA; and (c) what are students views of CITA. This paper is destined for publication, but at the time of the publication of this thesis, it is in pre-print format. The manuscript opens with an overview of the CITA system, the theoretical framework, the structure of the tutorials, and the schedule of development from 2015 to 2016. Then, I present a comparison of student learning between CITA users and non-CITA users from the three versions of the CITA system in the summer 2015, fall 2015, and spring 2016 semesters.

In order to get a full picture of the efficacy of the CITA system, students were assessed using a variety of methods that focused on different aspects of their physics education. First, I conducted a study to determine if certain groups of students were using the CITA system more than others. Then, I performed an analysis of performance on a variety of assignments between students who tended to use the system more versus those who used the system less. This analysis included overall course grades, exam scores, BEMA scores, and the assessment of a multi-step problem. I supplemented this study with a qualitative overview of student comments on the exit survey and focus group feedback.

4.2 Introduction

Due to the increased enrollment in foundational, university level physics courses, professors often find it difficult to provide personal assistance to students throughout the semester. In particular, weekly homework assignments require teachers to dedicate a large amount of time and energy to answer questions from students. This might include extra time after classes, weekly office hours, or public study sessions run by teaching assistants. When direct help from a physics faculty member is not available, students need a learning tool that guides them through homework problems - especially in an environment where non-traditional students with diverse backgrounds attend class [1–3].

Faculty cannot realistically be available to students at any time of the day. Thus, our research group wished to develop a tutorial system that could provide basic guidance to students on their homework and support their learning of physics. This project is called the Computerized Interactive Teaching Assistant (CITA). The CITA program strives to guide students through their homework and support students in learning some of the nuanced problem solving techniques in physics.

This paper reports the results of the CITA system thus far. Specifically, the research group (henceforth referred to as "we") wishes to answer the following research questions:

1. **Who uses CITA?** Do all students use CITA equally as much as their colleagues? Or do certain groups of students prefer using CITA over other resources?

2. **Does CITA improve student performance in class?** Are there significant differences in performance between students who use CITA and those who do not? Do all students benefit from the CITA system, regardless of background?

3. **What are students views of CITA?** Are students satisfied with system? Do they believe that it is helpful to their learning of physics? How does it compare to other resources?

4.3 Description of the Class

The development of the CITA system took place in the calculus-based, introductory E&M course at Purdue University. The second semester of introductory physics is called Electricity and Optics (coded as PHYS 24100 or PHYS 24100D). This course is targeted toward sophomore engineering majors who are about to start taking concentration-specific classes. Electrical engineering majors are the one exception because they take a different introductory electromagnetism course [9]. Some of the major topics covered in Electricity and Optics are electric charge, electromagnetic fields, Maxwell's equations, geometric optics, and interference effects.

The "D" in PHYS 24100D signifies that a student has signed up for a distance learning section in the fall or spring semester (i.e. an online section). This code is not necessary during the summer semesters because the course is exclusively offered online. The content in the on-campus and online sections of the course is exactly the same; all students, regardless of section, use *Physics for Scientists and Engineers* as the primary textbook [10], take weekly recitation quizzes, complete followup lecture quizzes, and attend the same exams. They differ in the fact that the online sections watch video lectures and attend online recitations through Cisco WebEx [11] while the on-campus sections attend lectures and recitations on campus. Additionally, online courses take their lecture and recitation quizzes online through their online homework system while on-campus students take their quizzes during their scheduled lectures and recitations.

Most engineering students take PHYS 24100 in the fall semester (called the "on-semester"). Alternatively, some engineering students choose to take the course in the spring or summer (called the "off-semesters") for a variety of reasons: co-op work schedule, transferred schools, repeating the course, etc. Table 4.1 shows the total enrollment in the campus and online sections of Electricity and Optics over the last three years. During each of the fall and spring semesters, just under 10% of the total class is enrolled in PHYS 24100D. Summer enrollment has risen steadily since the online version of the course was created.

Table 4.1
Enrollment in the campus and online sections of Electricity and Optics from spring 2014 through summer 2016. All of the enrollment numbers ignore dropped students.

Semester	Online	Campus	Percent Online
Spring 2014	41	465	8.1%
Summer 2014	153	0	100.0%
Fall 2014	62	725	7.9%
Spring 2015	43	474	8.3%
Summer 2015	187	0	100.0%
Fall 2015	69	728	8.7%
Spring 2016	36	339	9.6%
Summer 2016	222	0	100.0%

Electricity and Optics is generally taken after students complete PHYS 17200 (introductory mechanics) or if a student has earned AP credit to skip that course. PHYS 17200 is currently taught using the first semester of the Matter and Interactions curriculum by Ruth Chabay and Bruce Sherwood [12].

4.4 Literature Review

In order to develop and analyze a successful tutorial system, one must start by identifying the theoretical basis for how the system will teach. Thus, we started with our fundamental learning theory and build outwards from there. A learning theory (or learning philosophy) is a theoretical framework that describes how people capture, process, retain, and interpret information. The learning theory that informed our work is constructivism; this idea has been developed over many years by a number of researchers [112–114].

Whenever a student learns something new, he or she always has some prior knowledge of the subject. This knowledge can come from their previous schooling (e.g. a formal science class) or from other life experiences (e.g. piecing together the concept of acceleration after riding on a roller coaster). A student's knowledge includes both normative and non-normative models of how the world works. Constructivism asserts that learning is an active process by which new information is linked to a learner's previous models of how the world works. Learning is not a passive process [21,22,115].

Constructivist theory is framed in several ways; the most common strands of constructivism are: cognitive constructivism, radical constructivism, and social constructivism. We employed a framework based on cognitive constructivism, which asserts that the acquisition of knowledge is an adaptive process that results from active model building by an individual learner. Therefore, learning is the result of adjusting mental models to more accurately reflect external reality.

Constructivism is very applicable in the planning of lessons and design of tutorials. A constructivist would argue that learning is the process of generating and revising models as a person is subjected to a new experience. Thus, a student in a physics class is actively building and revising models based on the course material and their pre-existing models of the world. Thus, instructors should design a learning experience that helps students generate a fundamental understanding of the material rather than simply encourage memorization. Effective teachers take into consideration their students' pre-existing knowledge and plan lessons accordingly. They must be especially interested in the errors made by students because those errors are the points at which true learning can begin [25].

In a physics class, instructors are interested in more than just building models; instructors want students to be able to apply those models to solve real-world problems. The modern theory of problem solving is rooted in the work of George Polya. He described a deceptively simple strategy for solving a problem. First, one must fully understand the problem, including all of the qualitative concepts and assumptions that go along with it. Second, one must devise a plan to solve the problem. Third, one must execute the plan in a controlled way. Finally, one must check the final answer to ensure that it agrees with the initial understanding of the problem. Although this seems completely trivial, most problem-solvers do not make it past the first step [39].

Further research has been done to determine how students solve problems in physics classes. Many of these studies make a distinction between “master” and “novice” problem solvers (based on the pioneering work by Adriann DeGroot [116]). A novice problem solver does not use a structured method to work through a problem. Instead, he or she works in an erratic way. This might include identifying similar diagrams, looking through a textbook for the “right equation”, or looking for similar solutions online [41, 42]. Redish et al. identified a number of “epistemic games” that students play to work through problems in physics [43]. For example, he described how novice problem solvers often work backwards - skipping the initial analysis described by Polya and instead fitting random equations to known quantities.

Conversely, master problem solvers have discipline and a well-defined strategy while problem solving; they can carefully approach, dissect, solve, and check a problem. They start from first principles and work their way towards an answer [45]. The difference between expert and novice problem solvers in physics can be framed in the context of constructivism; experts have a thorough understanding of the laws of physics and feel comfortable starting from first principles. Novices do not have a good mental model of the laws of physics, so they need some sort of example by which they can relate the material to something they already know.

It should be noted that the classifications of master and novice are not problem specific. Rather, they describe the mindset and thought process used by those who can break down an arbitrary problem into smaller parts and those that cannot. Introductory physics students (largely novice problem solvers) tend to solve problems by memorization and surface features. However, they can grow into expert problem solvers by the use of educational scaffolding. Scaffolding provides an environment where students can systematically learn concepts, all the while gaining independence [48].

The idea of educational scaffolding is based on the work of Lev Vygotsky. He proposed that young children can accomplish sophisticated tasks that were normally outside of their reach with the assistance of an adult. The adult creates a series of small steps that the child can follow with relative ease. Then, the child can connect each step to arrive at a final solution. The area between what is known and what is not known is called the Zone of Proximal Development (ZPD) [30, 117]. In order to maximize learning, one must keep the child in the ZPD.

Within the context of education research, scaffolding refers to a variety of techniques that systematically move a student towards a deep understanding of the material (as dictated by the constructivist philosophy). Teachers provide successive levels of temporary support that students can use to structure their learning. These levels of support are removed when the student no longer needs them. In this research project the CITA tutorials provided the scaffolding needed to learn the physics concepts and problem solving skills. The content of the tutorials came from the course textbook and notes. Mayer's design principles of multimedia learning dictated the look and structure of CITA [59].

4.5 Design of the CITA System

4.5.1 Structure of the Tutorials

The CITA tutorial system is built on top of CHIP (Computerized Homework in Physics). CHIP is based on an online homework system called CPlite which was developed at the University of Illinois in Urbana-Champaign (UIUC). It has been under constant extension and revision since its adoption at Purdue University. The research group decided to expand the CHIP system rather than build a new system for a few reasons. First, we could continue to use the existing database of physics questions available in CHIP, including the "interactive examples" (i.e. interactive tutorials) developed by the faculty at UIUC for CPlite. Second, we could build on the existing CHIP infrastructure of problem presentation, feedback, scoring, and recording of student inputs. Third, we could easily compare grade records across semesters, both before and after CITA. Finally, there is a dedicated support staff at the university that has experience handling questions and error reports by the students.

CITA is designed to be a sequence of successively more detailed tutorials that target students with a variety of skill levels. Our early qualitative analysis of the student body indicated that students desire focused help right when they encounter a problem. Otherwise, they will turn to other sources like Google or Yahoo Answers (which will detract from the learning process). Thus, the project has two key ideas:

1. Provide students with focused feedback to any incorrect answers.

2. Give students the power to easily find and traverse a tutorial that fits their skill level.

In order to build this tiered system, we divided up the structure of CITA into three main parts: Shallow CITA, Immersive CITA, and Postscripts. None of these parts are meant to work independently of the other. Rather, these three items are built to work together to take a student through Polya's problem solving process in a controlled way.

An example of Shallow CITA is shown in Figure 4.1. A student is given a homework problem that must be solved for a grade. If the student makes an easily identifiable mistake on the problem (e.g. a unit error, a sign error, a simple conceptual error) a red box will pop up and offer specific feedback on how he or she should correct the mistake to proceed. Shallow CITA assumes that a student generally understands the major concepts in the problem and is close to a correct solution. If Shallow CITA cannot identify the mistake made by the student, it suggests that he or she attempt an Immersive CITA tutorial and provides a button.

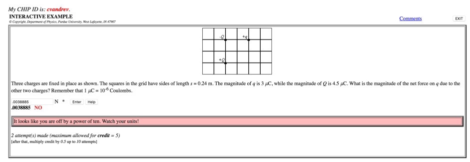

Figure 4.1. A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error.

Immersive CITA was built to provide detailed tutorials about the specific concepts and strategies in a given homework problem (see Figure 4.2). During the fall semester of 2015, we paired step-by-step tutorials with 109 of the problems on the homework (just under 80% of the total problem count). We decided to omit tutorials from problems that consisted of more than three graded parts (since the multiple parts already hinted at a path to the solution) and true/false questions. Additionally, problems within a homework assignment that repeated similar concepts were given truncated tutorials that outlined the solution to the problem in a few sentences and/or referred the student to a relevant Immersive CITA tutorial in a similar problem.

During the spring semester of 2016, we updated most of the truncated tutorials mentioned above to detailed step-by-step tutorials (even if it meant there was repetition of tutorials within a single homework assignment). Additionally, we added a filter to 74 of the Immersive CITA tutorials (slightly over 50%). We asked students to rate their confidence with the problem and decide if they wanted a detailed, step-by-step tutorial or a faster, general overview of the solution. Depending on their answer, the system would filter the student body into one path or the other. Wherever possible, we tried to keep the detailed step-by-step tutorials the same between the fall 2015 and spring 2016 semesters so that we could make consistent comparisons. Again, about 80% of the homework problems included some form of tutorial.

As much as possible, we wanted students to work through the solution of the problem rather than just read a transcript outlining the solution. Thus, each step of

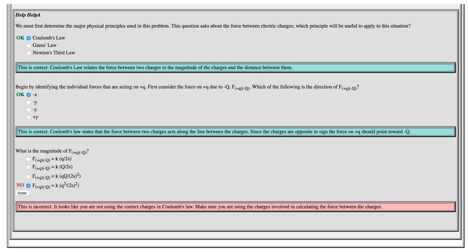

Figure 4.2. A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion.

an Immersive CITA tutorial consists of just a few sentences, followed by a question. In this way, the solution is methodically completed by the student. Each Immersive CITA tutorial followed a similar structure based on Polya's problem solving strategy. First, we spent a few steps describing some of the major conceptual ideas contained in the solution (studies have shown that developing the conceptual ideas of a problem improve the later quantitative analysis [51,52]). Next, the tutorial devised and worked through the solution, referring to lecture slides, class notes, and previous steps of the analysis.

Students could enter and exit an Immersive CITA tutorial at any time. Thus, the final step of Polya's problem solving process was separated from Immersive CITA and placed in what we call the Postscript. This ensured that all students received appropriate closing comments after a problem was solved such as a review of the methods used, a description of some common pitfalls in the analysis, or the relation of the problem to other problems in their homework.

Figure 4.3 shows the Postscript. After a student correctly answers a homework question, the Poscript reviews Polya's final step in the problem solving process. We ask students to think about the work that they just completed by offering simple variations of the problem. For example, if a student determined the force on a point charge was equal to zero, a postscript might ask what would happen if the charge were perturbed from its current location slightly. During the spring of 2016, each of the postscripts included a "Real World Application" - an example where the setup, model, or solution of the problem can be found in the realm of engineering.

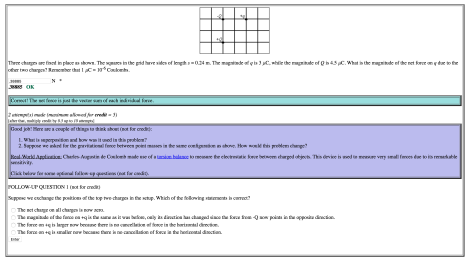

Figure 4.3. A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer.

4.5.2 Schedule of Development

Development of the CITA system began in the spring semester of 2015. Over the next year and a half, we used a design-based research approach to build, assess, and fine-tune the tutorials (see Section "Qualitative and Quantitative Connections"). The schedule of development is shown in Table 4.2.

Version 1 of the system was our first foray into development. During this time, we proceeded slowly, designing the structure that would be used in the branching system and ensuring that the tutorials worked for students. We were able to provide CITA tutorials for approximately 50% of the homework problems (while the remaining problems contained no help of any kind). The original "interactive examples" that were developed by UIUC were given Shallow CITA feedback and are included in the statistic above.

Version 2 of the system contained the step-by-step tutorials as described in the previous section. This version of the system was designed to serve two purposes. First, it was an intermediary step between the initial and final system so that we could break the development into two roughly equal parts. Second, it allowed us to run an analysis of a more traditional tutorial system in order to get some baseline data on how it works. During the fall semester of 2015, there was a problem in development, so the online courses were given Version 1 of the system while the on-campus classes received Version 2. This has been taken into account within the analysis.

Version 3 of the system consists of problems that offer the choice between two different levels of scaffolding (as described in the previous sections). This is the final goal of this particular research project. Of course, future researchers will take the system in new directions.

Table 4.2
The schedule of development of the CITA system. In the spring semester of 2015 and prior, the CHIP homework system included 29 "interactive examples" that were originally developed by the faculty at UIUC (out of 139 homework problems total). These 29 problems were paired with step-by-step tutorials and made up approximately 21% of the homework problems seen by students.

Semester	**Section**	**Version**
Spring 2014	Campus	None
	Online	None
Summer 2014	Online	None
Fall 2014	Campus	None
	Online	None
Spring 2015	Campus	None
	Online	None
Summer 2015	Online	1
Fall 2015	Campus	2
	Online	1
Spring 2016	Campus	3
	Online	3
Summer 2016	Online	3

4.6 Methods of Analysis

In this study we employed mixed-methods techniques to determine how the different tutorial structures affect students' performance in each semester. We outline our procedures below.

4.6.1 Quantitative Procedures

For the quantitative analysis, we used linear models, ANOVA analysis, and chi-squared tests where appropriate. The traditional cutoff of $p = 0.05$ was used for tests of statistical significance along with Cohen's guidelines for interpreting effect sizes [95]. Measures of effect size included Cramer's V, eta squared, and R^2 (where appropriate) [97, 98].

4.6.2 Quantitative Data Sources

Exam Scores

Due to the inclusion of the tutorials, homework scores predictably increased between spring 2015 (pre-CITA) and spring 2016 (post-CITA). However, these scores do not tell us how students perform when the scaffolding is removed. Thus, we seeded the midterm and final exams with questions from previous semesters so that we could determine if the scaffolding improved student problem solving skills. Our control groups for the exams were the spring 2014, summer 2014, fall 2014, and spring 2015 semesters. These classes were compared to the summer 2015, fall 2015, spring 2016, and summer 2016 semesters. Generally, a single control semester was compared to a single test semester; for a few problems, we could compare three semesters instead of two.

Brief Electricity and Magnetism Assessment

The BEMA was developed by Ruth Chabay, Bruce Sherwood, and Fred Reif in 1997 [100]. Although it was originally designed to measure student retention of electricity and magnetism concepts three months to five semesters after completing an introductory electricity and magnetism course, it is now often used to analyze student learning between the beginning and end of the semester. It is a useful tool to assess the understanding of electricity and magnetism concepts that are covered in a college-level calculus-based introductory physics course.

The BEMA is a multiple choice test consisting of qualitative questions and a few simple calculations. Lin Ding et al. performed an analysis of the BEMA, showing that it is a reliable assessment tool for introductory electricity and magnetism courses [101, 102]. Thus, we will use the BEMA exam to assess student conceptual understanding of introductory electromagnetism topics. Our control group for the BEMA is the spring 2015 semester; this class will be compared to the summer 2015, fall 2015, spring 2016, and summer 2016 semesters.

Multi-Step Problem

Chabay and Sherwood demonstrated a technique to analyze how well students can solve a non-trivial, multi-step problem [103,104]. They gave their students a complex problem and tracked how far students made it into the analysis before getting stuck. Then, they plotted the curves for how many students made it to certain points in the problem before getting stuck.

We administered a non-trivial problem to on-campus students during their recitations and graded their responses with a standardized rubric (see Appendix 4.12). Students were first given a 20 minute review of the major topics in the problem by their teaching assistant. Then, the problem was given under test-like conditions (i.e. no books, no notes, no collaboration with others). Once again, this allowed us to assess how students fared without the scaffolding provided by CITA. The control group for this analysis was the spring 2015 on-campus students while the experimental groups were the fall 2015 and spring 2016 on-campus students.

Student Exit Survey

Capping off our quantitative analysis was an online survey (based on a Likert Scale) given to students at the end of the semester. It asked about their opinions of the CITA system as well as how they used it. This survey was required for a recitation quiz grade (as mandated by the internal review board at Purdue University). However, students had the option to select “I Prefer Not to Disclose” for each question. The purpose of this survey was to learn about how students felt about the CITA system in terms of effectiveness, workload, and communication. See Appendix 4.13 for a copy of the student exit survey.

4.6.3 Qualitative Procedures

Strategy of Inquiry

The qualitative part of this study analyzed student opinions about the CITA program as well as their general study habits. Qualitative data came from three main sources; first, the final question of the exit survey provided an open space for students to make comments on the CITA system. Next, we analyzed student comments on the Piazza help system which is used for online office hours. Finally, we offered students the opportunity to participate in a focus group session at the end of the semester to share some of their opinions. After the second exam (when we started on the AC circuits unit) we sent an email out to students asking for volunteers for a one hour focus group session. The sessions took place at the student’s convenience in the two weeks before their final exam. After each session, we rewarded participants with a ten dollar gift card.

Our strategy for analyzing the transcripts from the three sources above consisted of a three-cycle coding plan (Saldana's reference provides the descriptions for each type of code listed below [105]). In the first cycle, each of the transcripts was analyzed separately. It consisted of rounds of attribute coding and descriptive coding in order to get a sense of the recurring comments made by students. The second cycle also treated the transcripts separately. It consisted of rounds of elaborative coding and pattern coding that were used to synthesize the results of the first cycle and identify the major themes in the transcripts. The final cycle of coding pooled all of the sources together. It consisted of evaluation coding and longitudinal coding to identify how comments compared between semesters.

Analysis Procedures

Once the research team compiled the transcripts from the focus group sessions, Piazza comments, and written survey answers, we carefully parsed the data to find patterns. We used inductive analysis and creative synthesis as the analysis and reporting strategy. Patton describes this strategy as starting from the specific details of the inquiry, slowly finding patterns in the data until a broad theory can be synthesized [107]. It allowed us to mesh together the opinions of the participants in the focus groups in a controlled way.

As described in the previous section, our qualitative analysis used a three cycle system. First, we performed a coding of the transcripts looking for specific key words and phrases that highlight student opinions. Next, we compiled these keywords into categories that described student opinions within a semester. Finally, we compared student opinions between the semesters.

Roles of the Researchers

In qualitative studies a researcher must acknowledge that his or her presence in the experiment may alter the findings. Thus, researchers must take time beforehand to determine how their presence might influence the data as well as how to minimize this artifact [108]. In this study the principle investigators were Mr. Cyrus Vandrevala, Dr. Lynn Bryan, Dr. Andrew Hirsch, Dr. Hisao Nakanishi, and Dr. Laura Pyrak-Nolte.

Mr. Cyrus Vandrevala has been a teaching assistant for PHYS 24100 and PHYS 24100D since the fall semester of 2011. Additionally, he has coordinated the course during summer sessions and helped develop the online sections. This included editing course videos and setting up the online Piazza classroom [109]. During the study, he held a part-time assignment answering questions on Piazza so that students never came face-to-face with him. Thus, students were not interviewed by any faculty member who assigned them a grade in the class.

Dr. Hisao Nakanishi is the "father" of CHIP at Purdue University because he is one of the primary developers of the system. Due to his experience with CHIP, Hisao Nakanishi developed and implemented the underlying structure upon which the branching tutorials were built. Although he is not directly affiliated with PHYS 24100 or PHYS 24100D, he addressed student questions sent through the CHIP help system that pertain to errors in the problems, bugs in the code base, or administration issues. Additionally, he helped review all of the content that was uploaded to the website including: homework problems, exams, quizzes, tutorials, surveys, and the BEMA.

Dr. Laura Pyrak-Nolte has taught PHYS 24100 and PHYS 24100D since 2005 and coordinated the course since fall 2011. As the course coordinator, she is in charge of creating the curriculum for the class, designing the exams, assigning final grades, and approving all new changes to the homework system. Additionally, she is the lecturer that is seen and heard in the online lecture videos shown to the PHYS 24100D class. Due to the fact that Laura Pyrak-Nolte is the coordinator for the class, she did not conduct any of the focus group sessions or interviews with the students as this would put unnecessary pressure on the participants.

Drs. Lynn Bryan and Andrew Hirsch provided much of the educational theory that supports this study. Neither were directly involved with PHYS 24100 or PHYS

24100D. It should be noted that Andrew Hirsch is one of the main instructors of PHYS 17200 - the introductory physics class that is a prerequisite for PHYS 24100 and PHYS 24100D.

4.6.4 Focus Group Sessions

Focus group sessions are ideal for this study for a variety of reasons. First, they are flexible; they can be used for exploratory, explanatory, and evaluative research. Focus groups create a large volume of data with a range of viewpoints from all of the participants. Furthermore, a skilled moderator can ensure that this data has limited researcher influence. In a one-on-one interview, the interviewer cannot help but insert some of his or her own views into the discussion due to the questions asked. However, a focus group can "veer off topic". Since the members of a group can discuss what they feel is important, they can illuminate important points that the moderator might miss [106].

We used purposeful sampling as our design strategy. Patton defines this strategy as one where the researcher chooses specific participants for the study in order to gain insight about a specific phenomenon. He describes how cases are selected because they are "information rich" [107]. When students signed up to volunteer for our focus group sessions, we asked them what their overall impression of the homework system was - "Generally Positive", "Generally Negative", or "Neutral". We then divided the participants into groups of four based on their reported overall impression of the system and whether they were in an online or on-campus section. We wanted to match similar opinions beforehand so that students were comfortable sharing their opinions - whether they were positive or negative.

4.7 Results

4.7.1 Usage of CITA

Students were not required to use any of the tutorials or answer any of the follow-up questions in the homework system. Rather, they were given the recommendation through Shallow CITA to traverse the full tutorial if the system could not recognize their answer. Thus, we start off with an analysis of who used the system as a function of demographic information. In each case, we divided the classes into online and on-campus sections as mentioned above.

Usage by Semester

During each semester, we collected two types of usage data. First, we counted the actual number of clicks that students made within the homework problems. Second, we asked students to report their usage of CITA through an exit survey at the end of the semester (see Appendix 4.13). We wanted to see if student interaction with the system increased as the system grew in sophistication.

Figure 4.4 shows histograms of the total number of clicks recorded for all homework problems in each semester by each student. These clicks include attempts at the graded portion of the problem as well as steps within the Immersive CITA tutorials. We might expect that as more problems are paired with tutorials, the average number of clicks recorded in a semester will increase. ANOVA analysis shows that the change in the mean number of clicks in the on-campus sections is statistically significant with a small effect size (eta squared) of 0.022. However, the change in the mean number of clicks in the online sections is not statistically significant.

Surprisingly, the number of clicks in the spring 2016 semester was not appreciably different from the fall 2015 semester in the on-campus sections. Additionally, we must accept the null hypothesis that the change in the mean number of clicks was not statistically significant in the online sections of Electricity and Optics.

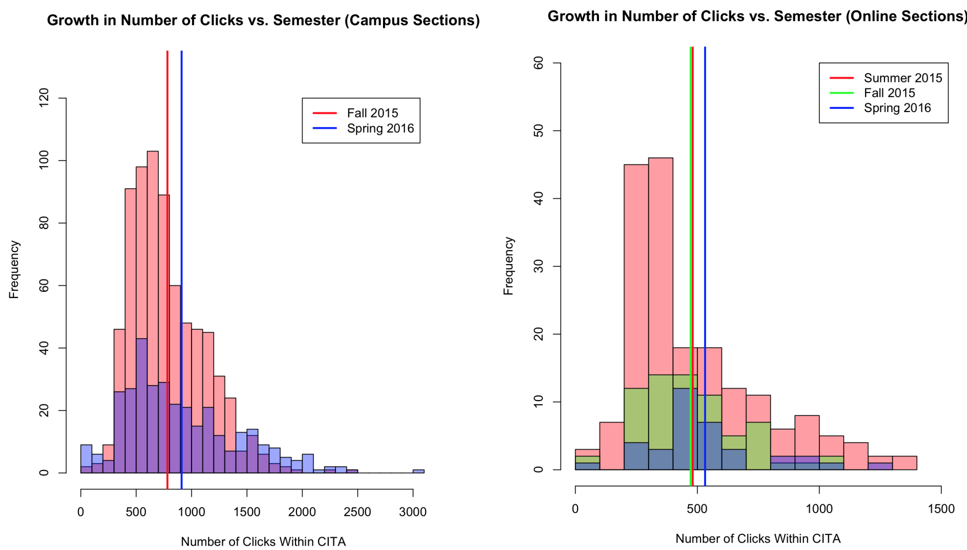

Figure 4.4. The growth in the number of clicks between the different semesters of PHYS 24100. The x-axes give the number of clicks while the y-axes give the number of students. There were 138 homework questions given in each semester.

Further analysis of student answers on the exit survey confirms this result. Both questions one and two of the exit survey showed no statistically significant difference in mean score between the summer 2015, fall 2015, and spring 2016 semesters. Students were using the CITA system, but there was little to no growth in the use of the system between the fall 2015 and spring 2016 semesters.

It should be noted that the average number of clicks between the online and on-campus sections within a semester is very different. On-campus students in each semester tend to use the tutorial system more than their online counterparts with a medium effect size (eta squared) of 0.057 in the fall of 2015 and 0.045 in the spring of 2016.

Usage by Gender

A few studies have shown that there is a difference in the way that males and females use online homework systems and social networks [118–120]. Thus, we studied the usage of CITA between the two groups. Table 4.3 shows the p-values for the ANOVA analysis comparing the difference in the mean number of clicks between males and females in each semester and each section. None of the sections or semesters show a statistically significant difference between males and females in recorded the number of clicks.

Table 4.3
P-values of the ANOVA analysis of the number of clicks between males and females in different sections and semesters.

	Campus Sections	**Online Sections**
Summer 2015	NA	0.56
Fall 2015	0.06	0.65
Spring 2016	0.80	0.56

Table 4.4 shows the p-values for the chi-squared tests comparing the difference in the answers by males and females on the first question of the exit survey in each semester and each section. Again, none of the sections or semesters show a statistically significant difference between males and females in the number of clicks.

Table 4.4
P-values of the chi-squared analysis of the number of clicks between males and females in different sections and semesters.

	Campus Sections	**Online Sections**
Summer 2015	NA	0.85
Fall 2015	0.19	0.70
Spring 2016	0.08	0.51

Although a few of the on-campus sections revealed p-values close to the 0.05 cutoff, we cannot reject the null hypothesis that the male and female populations are different. Unlike the results of the studies from above, our male and female student populations seem to engage with the CITA system approximately equally.

Usage by Previous Performance

We wanted to see if previous performance in an introductory physics class affected student's use of the CITA system. Table 4.5 shows the responses to question one of the exit survey as a function of previous grade in PHYS 17200 and semester. Chi squared tests of the data show that there were no statistically significant differences in the self-reported use of CITA in the summer 2015 and spring 2016 semesters. However, the fall 2015 semester showed a statistically significant difference in the populations (when the D grade data was omitted) with $p = 0.01$ and an effect size (Cramer's V) of 0.14.

In general, students who scored an A or a B in PHYS 17200 tended to use the tutorial more during the fall 2015 semester. We also filtered the student populations by self-reported previous performance in their introductory calculus class, but found no statistically significant differences in the grade levels. With the exception of the fall 2015 break by PHYS 17200 grade levels, students report using the system at the same level, no matter their skill level in physics and calculus.

Preference for Using CITA

Our qualitative analysis adds depth to the results from above. When asked what resources they use to complete the homework, we received a wealth of responses from the participants. Each student adopted a set of tools that helped him or her finish the homework as quickly and efficiently as possible. Although every student is different, there are some strong patterns in the tools that are used by the class. Online sources are extremely popular, and the most commonly cited ones are Chegg, Yahoo! Answers, Hyperphysics, and Google. Some students do attend office hours, but the vast majority do not, usually due to time constraints. Students also reference

Table 4.5
Exit survey results of students in PHYS 24100, broken down by semester and PHYS 17200 grade. The question stated “I used the interactive walkthroughs to help me with homework problems”.

Summer 2015					
Grade	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
A	4	2	2	0	0
B	12	15	8	3	1
C	6	20	9	1	2
D	1	3	1	1	0
Fall 2015					
Grade	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
A	56	53	11	7	2
B	89	87	28	10	5
C	38	37	21	16	5
D	0	0	0	0	0
Spring 2016					
Grade	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
A	20	16	3	2	1
B	48	28	9	5	2
C	29	19	10	6	5
D	0	0	0	0	0

the textbook quite often. However, they describe using the textbook as a source of worked examples.

Students certainly mentioned the CITA tutorials in their list of resources. Some ranked the tutorials very highly, while others only referred to them as a last resort. In all of these cases, we found that the primary goal of a student in PHYS 24100 is to get the best grade on the homework as efficiently as possible. This means that any resource that is used must be focused and relevant to the individual student's problem at the given point in time. If the resource is not focused, students can easily find another source on the internet.

4.7.2 Student Improvement

Once we determined who was using the CITA system, we studied how the CITA tutorials affected students' understanding of physics and problem solving skills. We analyzed a variety of items including paired exam questions, the BEMA concept inventory, and a standardized multi-step problem. The results are shown below.

Overall Course Grades

One of the first steps in our analysis was to see if the CITA tutorials were helping all students, regardless of their initial study habits or experience. For example, what if the tutorials were showing artificially high results just because high-performing students were using them while low-performing students were largely ignoring them?

Early in the semester, we asked students to report their previous experiences in calculus and physics, as well as provide some demographic information (all students had the option to answer with "I Prefer Not to Disclose" should they choose). From this data, we divided the class into groups based on prior physics experience (Figure 4.5) and prior calculus experience (Figure 4.6) and plotted their overall class grade versus the number of clicks that they took on the CITA system. Due to the fact that we had very low number statistics in the online sections of PHYS 24100 during the fall 2015 and spring 2016 semesters, we combined the campus and online data points into one plot. The data from the summer of 2015 showed no clear trends and contained very large standard error bars. Thus, it is not included in the results below.

We see that students who used the CITA tutorials tended to have higher scores than their colleagues who did not. However, the improvement in course grades evened out after a certain number of clicks. Students who received an A in PHYS 17200 showed the least improvement from the tutorials, with the graph even taking a small downturn in the fall of 2015. Overall, the tutorials are a helpful tool for the students, but they can only supplement instruction.

We also cut the data sets by gender (Figure 4.7), domestic vs. foreign students (Figure 4.8), and section (Figure 4.9). With very few exceptions, we found that students in each group showed higher course performance with increased use of the CITA tutorials. This means that all students, regardless of previous physics experience, gender, or nationality do see some benefit from using the tutorials. One again, we see that students derive benefit from the tutorials up to a certain number of clicks. At that point, the supplementary instruction from the tutorials leads to very small learning gains.

The data from the fall semester can be fit to a linear model reasonably well. However, we found that the spring semester did not follow a linear growth, but instead tapered off after a certain number of clicks.

We then analyzed student performance in the course as a function of reported use of CITA. Figure 4.10 shows student grades in PHYS 24100 as a function of reported use of CITA. We see that in this particular case, the range of values between the different answers is very large. Thus, reporting that one uses CITA is not as indicative of performance as actually recording more clicks within the system.

Exam Grades

Although we were not able to administer exactly the same exams to each class in each semester, we made comparisons in two ways. First, we compared exam grades within a semester to see the difference in scores between students who used CITA versus those who did not use CITA as much. Second, we selected a few exam problems to appear on multiple final exams so that students could be compared across semesters. These problems covered topics from magnetic induction, AC circuits, properties of light, geometric optics, mirrors, and interference effects. Chi squared tests were used to compare the correct and incorrect scores from a given section (online and campus) between semesters. Comparisons were done between the CITA semesters (fall 2015 vs. spring 2016) as well as between pre-CITA and post-CITA semesters.

Table 4.6 shows the improvement of scores on test specific problems between fall 2015 and spring 2016. In the campus sections, two of the five tested questions showed improvement with a small effect size. The double slit question showed improvement, but it was just shy of the $p = 0.05$ cutoff to be considered statistically significant. The online sections showed no statistically significant growth from fall 2015 to spring 2016.

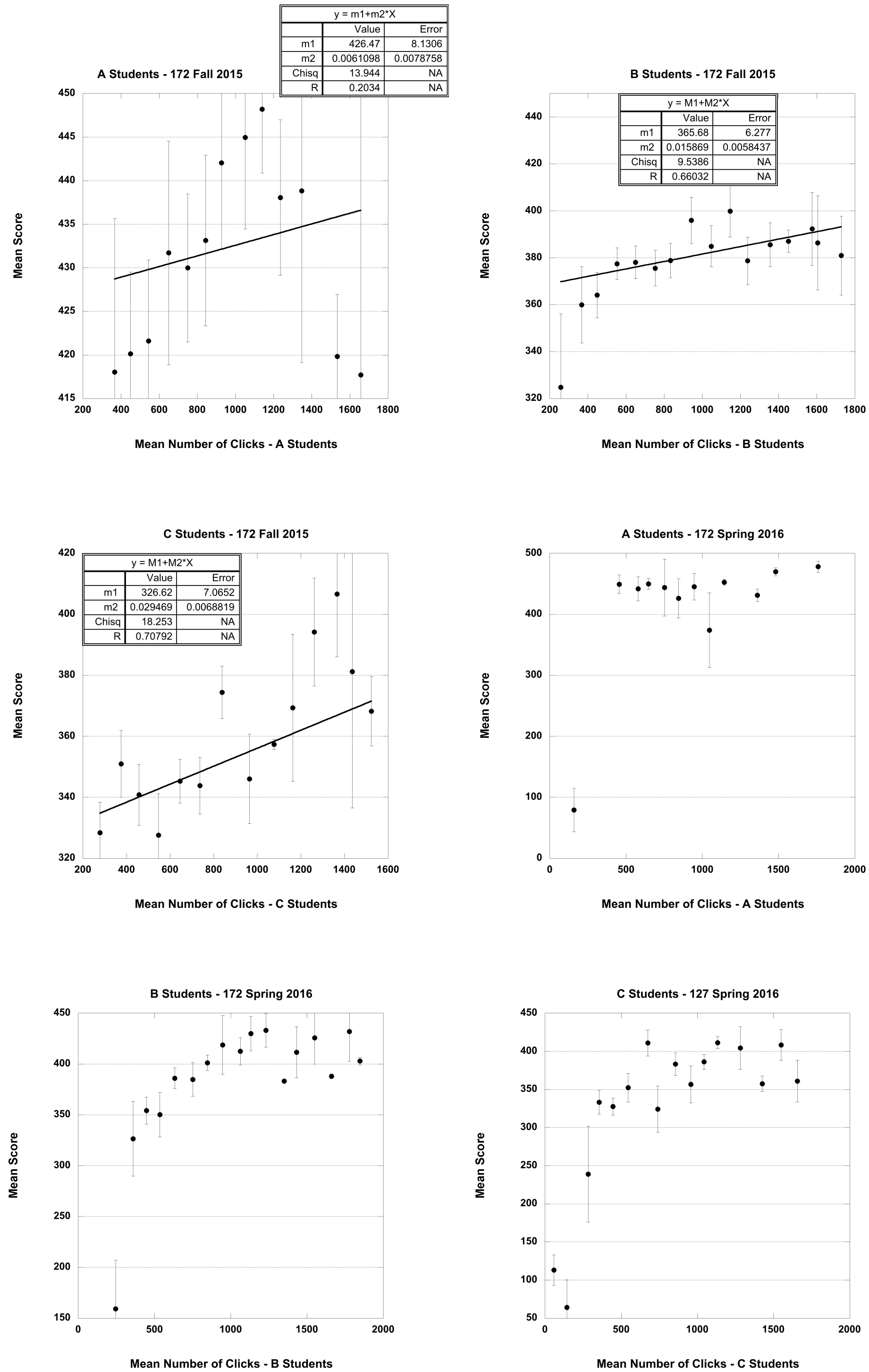

Figure 4.5. Overall course grade as a function of number of clicks within the CITA tutorial system. The students are binned by their previous grade in PHYS 17200 (i.e. introductory mechanics).

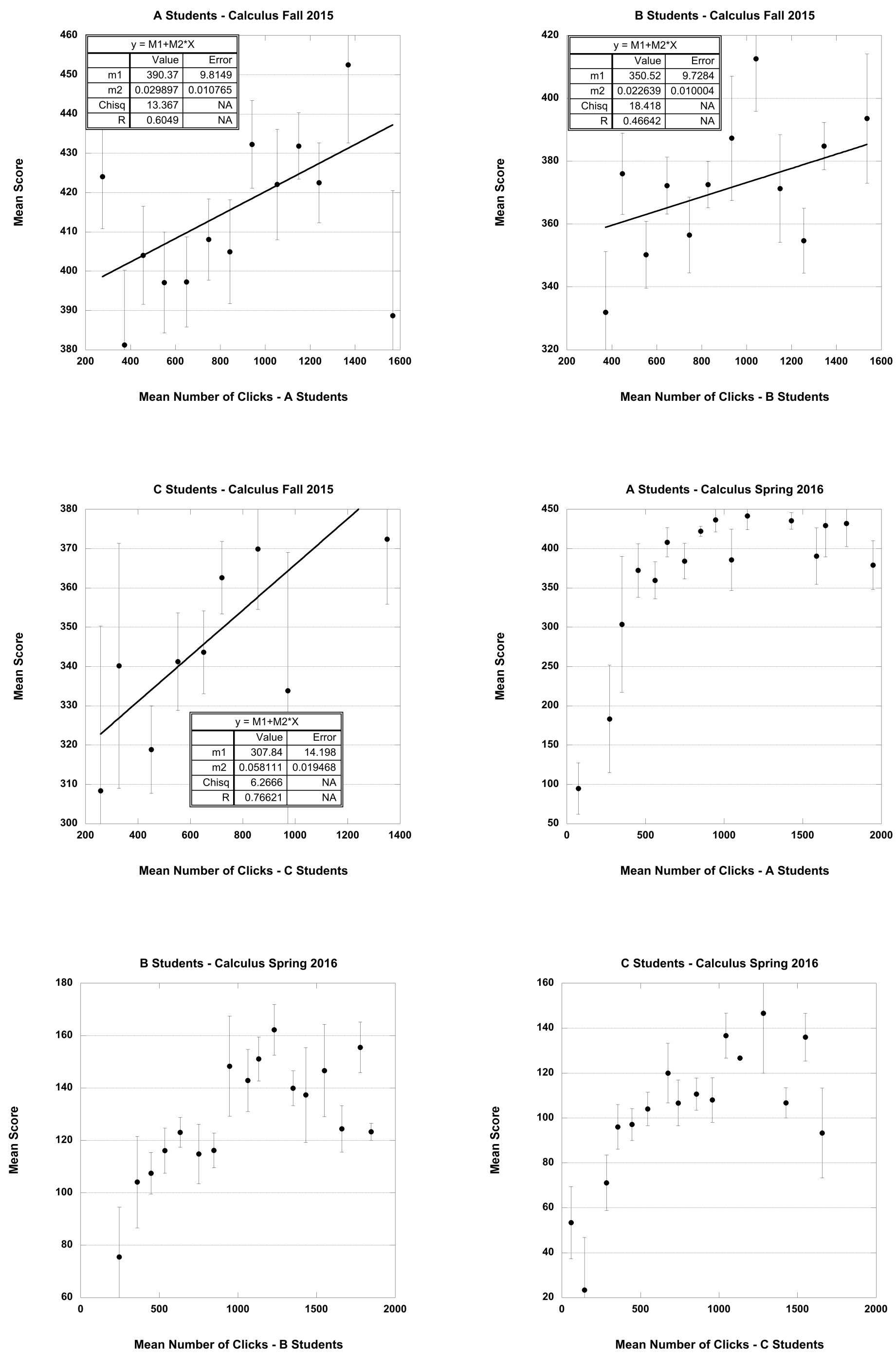

Figure 4.6. Overall course grade as a function of number of clicks within the CITA tutorial system. The students are binned by their previous grade in introductory calculus (i.e. Calculus 1).

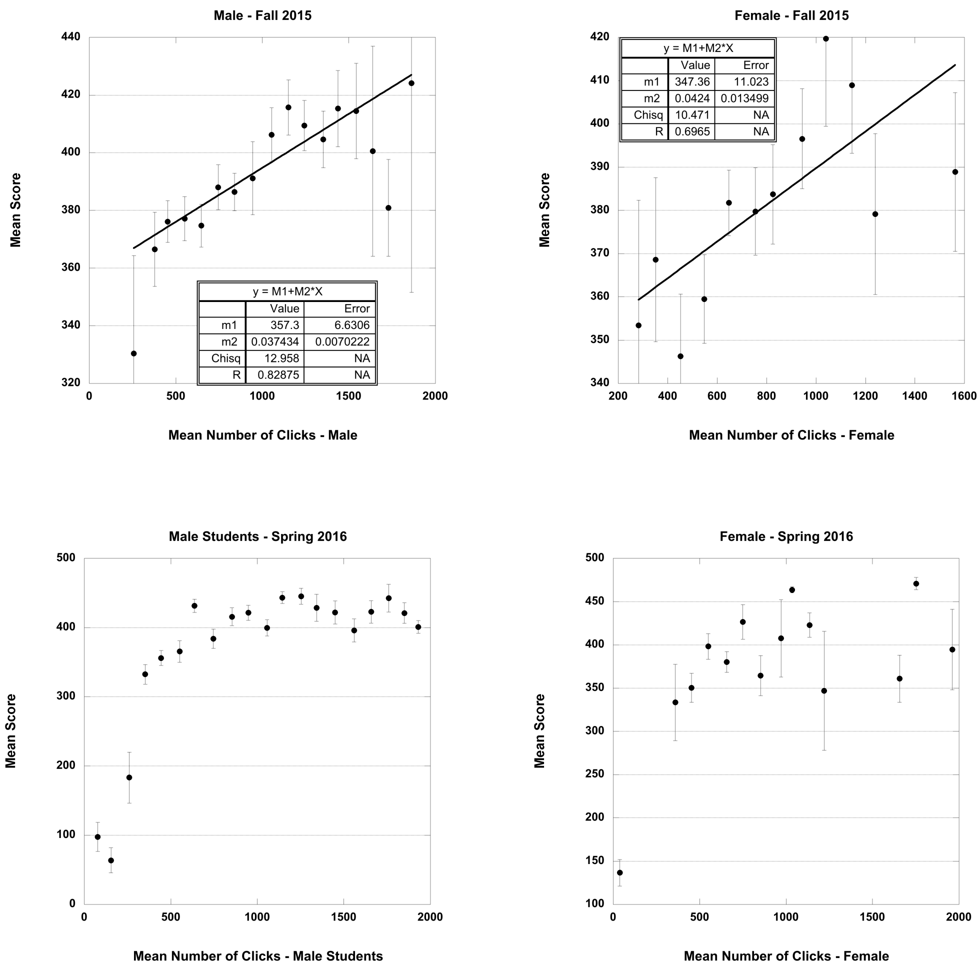

Figure 4.7. Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by gender.

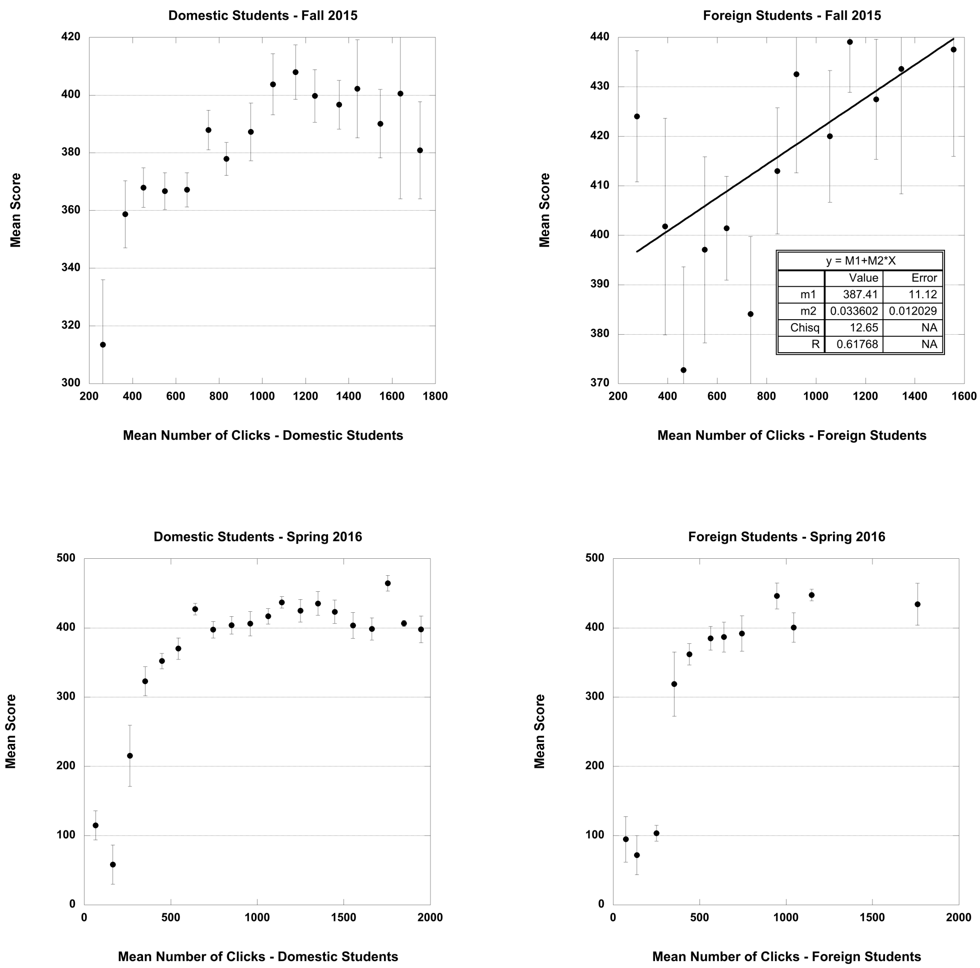

Figure 4.8. Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by domestic and foreign status.

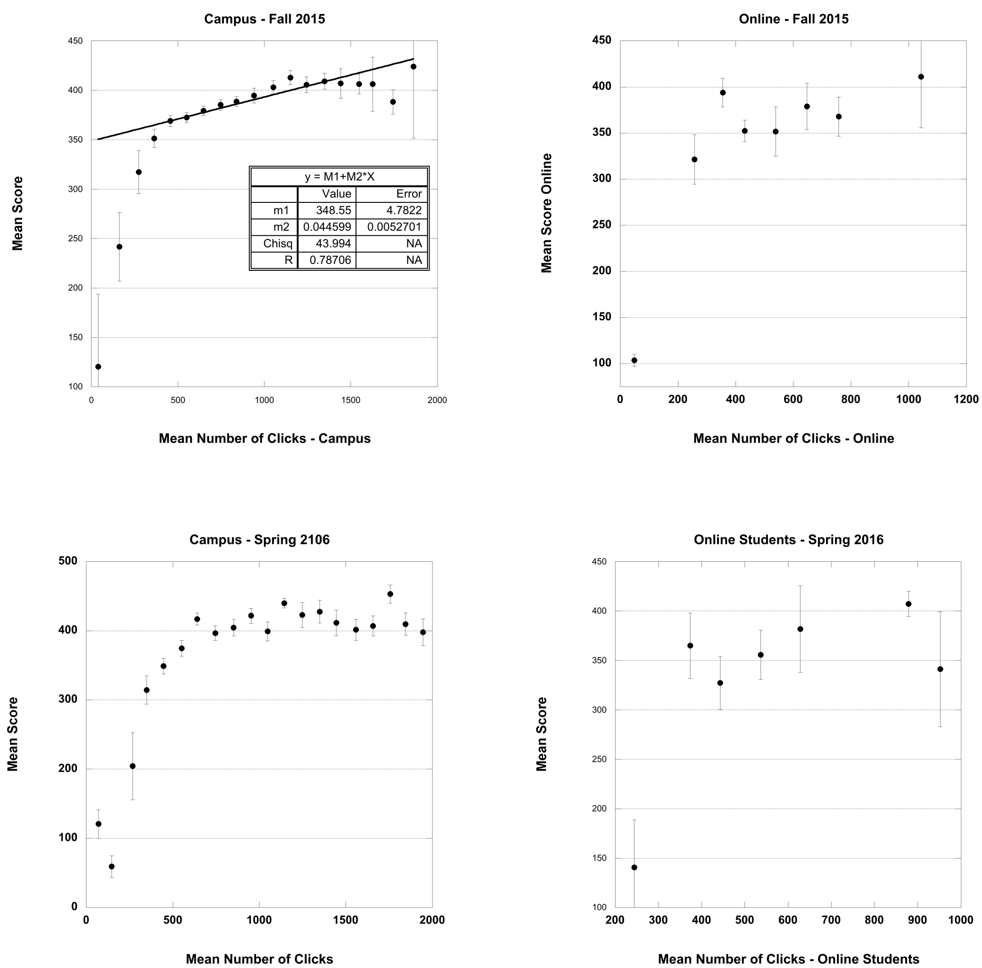

Figure 4.9. Overall course grade in PHYS 24100 as a function of number of clicks within the CITA tutorial system. The students are binned by their section (i.e. online vs. on-campus).

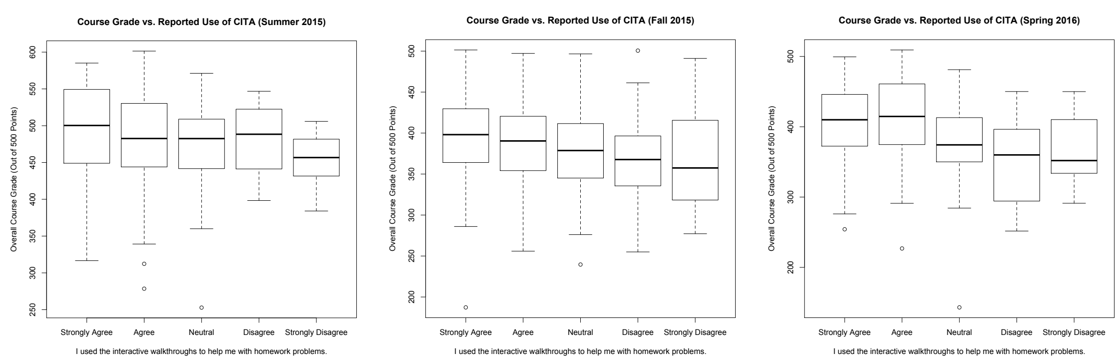

Figure 4.10. Overall course grade in PHYS 24100 compared with student answers on the exit survey.

Table 4.6
Chi-squared analysis of scores on selected test questions between the on-campus (top) and online (bottom) sections of the fall 2015 (fa15) and spring 2016 (sp16) semesters. We compare correct (C) and incorrect (I) answers between the semesters. The highlighted rows show statistically significant improvements between the fa15 and sp16.

Campus Sections						
Problem Description	**fa15 - C**	**fa15 - I**	**sp16 - C**	**sp16 - I**	**p-Value**	**Cramer's V**
Reflecting Sphere Floating in Laser Beam	110	604	69	248	0.0165	0.075
LC Oscillator Circuit	213	498	109	207	0.170	0.043
Metal Rod Moving on Two Rails	380	336	196	121	0.011	0.079
DC LR Circuit	448	268	211	106	0.246	0.036
Double Slit Apparatus, Slit Separation	581	135	271	44	0.069	0.057

Online Sections						
Problem Description	**fa15 - C**	**fa15 - I**	**sp16 - C**	**sp16 - I**	**p-Value**	**Cramer's V**
Reflecting Sphere Floating in Laser Beam	14	53	7	25	1.000	0.000
LC Oscillator Circuit	21	46	11	21	0.943	0.007
Metal Rod Moving on Two Rails	38	29	19	13	0.973	0.003
DC LR Circuit	42	25	17	15	0.492	0.069
Double Slit Apparatus, Slit Separation	48	51	16	2	0.475	0.079

The comparison of questions between pre-CITA and CITA semesters was not as conclusive. Although we did see a greater percent of students answer the exam questions correctly in CITA classes, none of the effects were statistically significant. This is in part due to small enrollments in the online sections in the fall 2015 and spring 2016 semesters, as well as the relatively low difficulty of the problems used. Table 4.7 shows the analysis of the questions between pre-CITA and CITA semesters.

Table 4.7
Chi-squared analysis of selected test questions between pre-CITA and CITA semesters of PHYS 24100. There are no statistically significant differences between the semesters (save one question).

Campus Sections

Problem Description	**First Semester**	**Second Semester**	**N**	**p-Value**	**Cramer's V**
Spherical Mirror (Real vs. Virtual Image)	Spring 2015	Spring 2016	757	0.216	0.045
Spherical Mirror (Focal Length)	Spring 2015	Spring 2016	757	207	0.027
Spherical Mirror (Shape of Mirror)	Spring 2015	Spring 2016	757	0.603	0.019

Online Sections

Problem Description	**First Semester**	**Second Semester**	**N**	**p-Value**	**Cramer's V**
Spherical Mirror (Real vs. Virtual Image)	Spring 2015	Spring 2016	69	0.182	0.161
Spherical Mirror (Focal Length)	Spring 2015	Spring 2016	69	0.182	0.161
Spherical Mirror (Shape of Mirror)	Spring 2015	Spring 2016	69	0.409	0.099
Light Moving Through a Prism	Summer 2015	Spring 2016	171	0.811	0.018
Properties of Light	Summer 2014	Summer 2015	297	0.034	0.151
Magnetic Field/Flux in a Solenoid	Summer 2014	Spring 2016	158	0.246	0.092
Image From a Dentist's Mirror	Summer 2014	Summer 2015	265	0.220	0.075
Diffraction Grating	Summer 2014	Summer 2015	105	0.846	0.019

Finally, we performed an analysis of exam grades within a semester versus student use of CITA. Figure 4.11 gives the total of the two midterm exams as a function of number of clicks on CITA. Again, we see that students who recorded more clicks in CITA generally performed better on the exams than their colleagues who recorded fewer clicks.

BEMA Concept Inventory

We administered the BEMA concept inventory to students at the beginning of the semester and after their second exam (i.e. after they had learned about magnetic induction and were moving onto AC circuits). From the raw score data, we calculated the gain of each student and compared the gains across the semesters. The gain is defined as the student's improvement in score over the total possible improvement for that student. Since the BEMA contains 31 items and we assigned each item an equal score, the normalized gain (G) becomes:

$$G = \frac{post - pre}{31 - pre} \tag{4.1}$$

where *post* and *pre* correspond to a students' post-test and pre-test scores. Similar to previous sections, we compared students' gains on the BEMA as a function of their actual use of the system and their reported use of the system. Figure 4.12 plots students' gains on the BEMA as a function of the number of clicks that they made on the system. The fall 2015 semester fits the linear model (p = 1.48e-5), while the spring 2016 does not fit as well due to the larger variance of the data points (p = 0.203). Although we cannot make any claims about the spring of 2016, we can say that an increasing number of clicks in the fall 2015 semester generally corresponds to an increasing gain on the BEMA.

Due to low numbers of students who took both the pre-test and post-test in the online sections in fall 2015 and spring 2016, we were not able to create reliable plots of the online sections. However, we can visualize the online sections in the summer

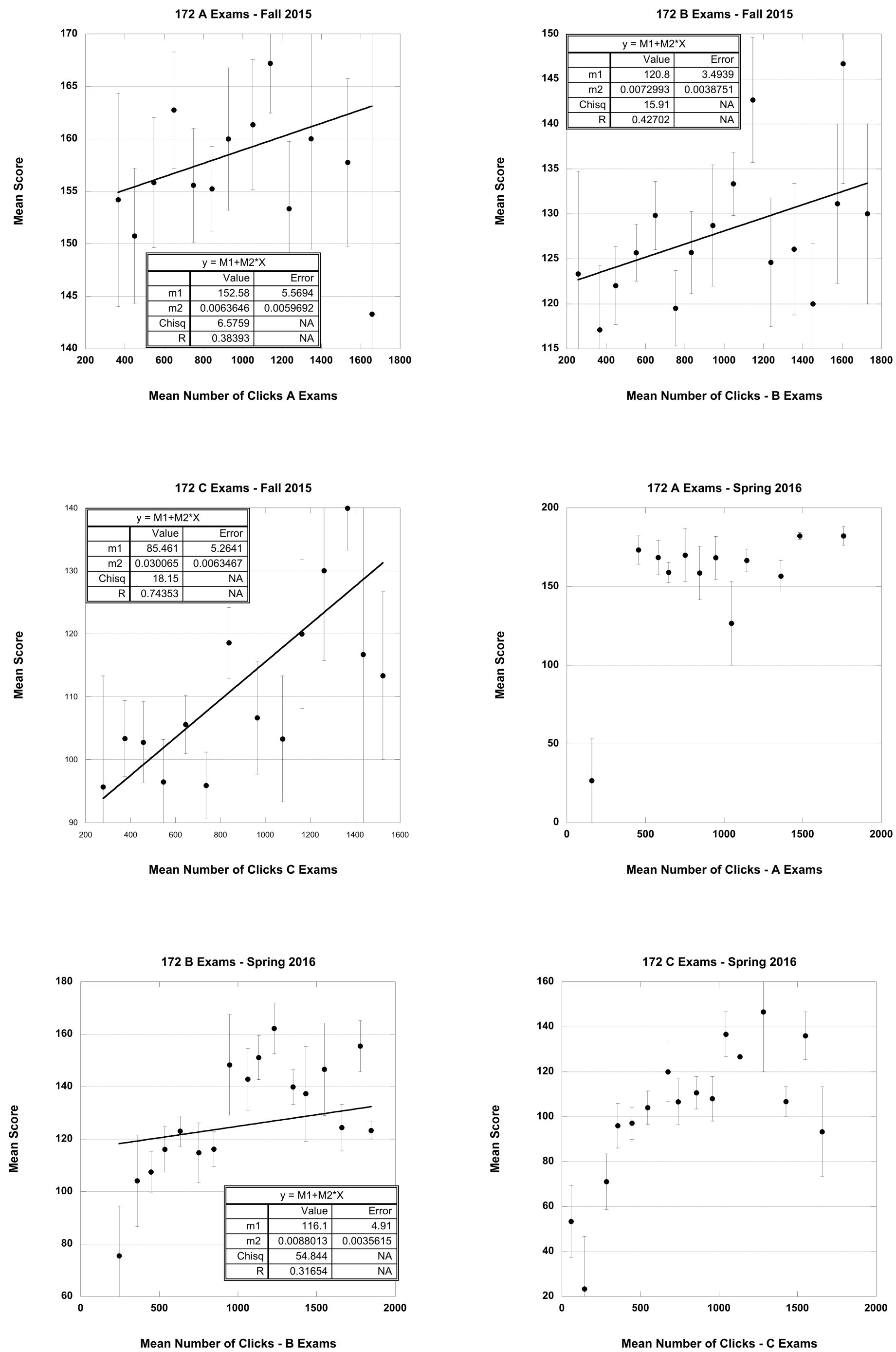

Figure 4.11. Total exam grade as a function of number of clicks within CITA.

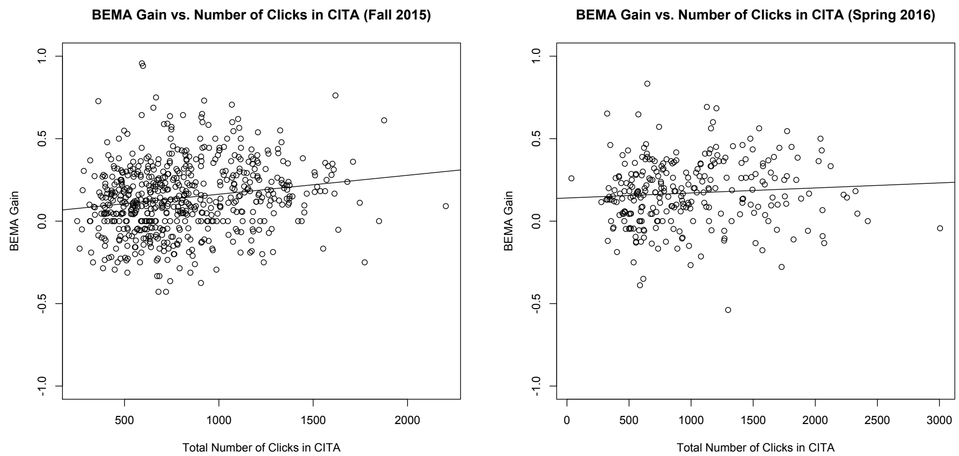

Figure 4.12. Student gains on the BEMA as a function of reported CITA use. The students shown in these plots are all in on-campus sections of PHYS 24100.

of 2015. Figure 4.13 plots the gain as a function of clicks in the summer of 2015. The plot on the left shows the unfiltered gains of all of the students in the semester. On first glance, this plot makes it look like the CITA tutorials have utterly failed students. However, it should be noted that we had a large problem with cheating on the online BEMA that semester, thus artificially inflating the scores. The plot on the right filters out pre-test and post-test scores above 15, thus showing a more accurate view of the class.

Next we compare BEMA gains to students' reported use of the CITA system. We performed an ANOVA analysis of the student gains, divided into groups by their answer to question one on the exit survey. Table 4.8 shows the results of the ANOVA analysis and Figure 4.14 shows a box plot of normalized gains when students are divided by their reported use of the CITA system. During the spring of 2016, we found that there was a statistically significant difference in BEMA gains when students reported greater use of the CITA system ("Strongly Agree") as compared to those that reported little use ("Strongly Disagree"). However, it should be noted that the range of the "Strongly Agree" group is much larger than the range of the "Strongly

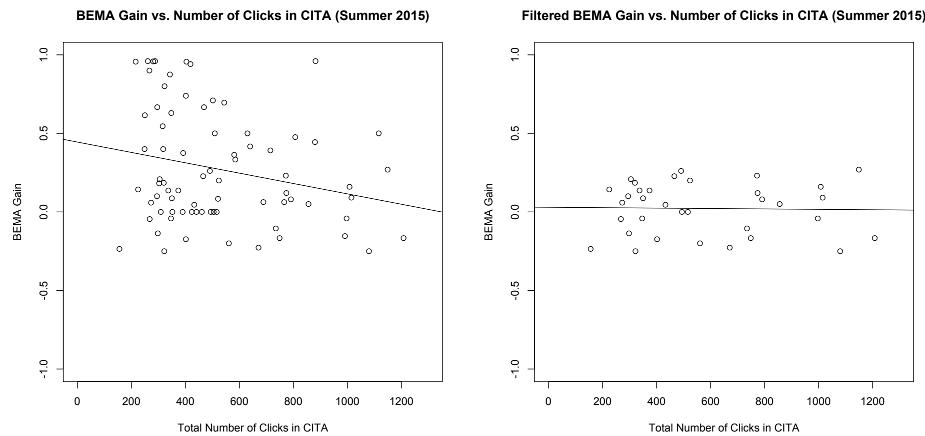

Figure 4.13. Student gains on the BEMA as a function of reported CITA use. The plot on the left shows the unfiltered online gains in the summer of 2015. The plot on the right filters out potential cheaters, thus giving a more accurate representation of the system.

Disagree" group, suggesting that a many students report using the system a lot, but their actual use does vary significantly.

Table 4.8
BEMA gains versus reported use of the CITA system for each semester. Spring 2016 shows the first statistically significant difference in gains between CITA users and non-users.

	Campus Sections			**Online Sections**		
Semester	**F-Value**	**P-Value**	**Eta Squared**	**F-Value**	**P-Value**	**Eta Squared**
Summer 2015	NA	NA	NA	0.471	0.757	0.027
Fall 2015	1.958	0.010	0.015	0.829	0.503	0.172
Spring 2016	3.305	0.012	0.053	0.001	0.975	0.001

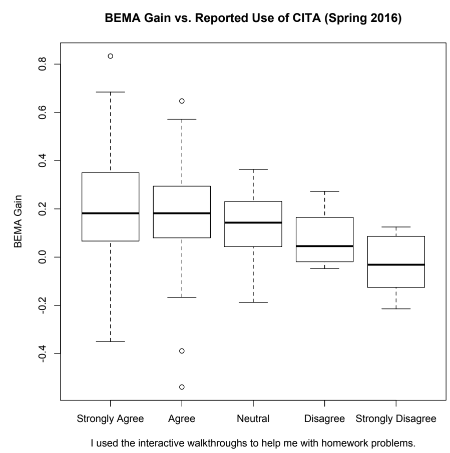

Figure 4.14. Student gains on the BEMA as a function of reported CITA use in the campus section of spring 2016.

Overall, we are seeing small gains for students in the campus sections of PHYS 24100 that use the CITA system. However, their online colleagues are not showing the same gains.

Multi-Step Problem

Our final analysis of student gains comes from the multi-step problem that we administered to the on-campus sections of PHYS 24100 during the spring 2015, fall 2015, and spring 2016 semesters. The breakdown of scores on the multi-step problem are shown in Table 4.9. Notice that the spring 2016 semester only has a total sample size of 198. This is because one of the teaching assistants accidentally kept a slide of equations up during the analysis, so the scores in his section were artificially inflated. Thus, we eliminated his sections from the analysis.

A chi-squared test of the data above yields a p-value of 0.001 and an effect size (Cramer's V) of 0.117. We are seeing distinct improvements in student's scores from semester to semester.

Additionally, we checked to see if students using CITA within a semester scored higher, on average, than students who did not use CITA during the semester. Surpris-

Table 4.9
Scores on the multi-step problem in each semester when it was administered, along with the average score and standard deviation of the semester.

Score	Spring 2015	Fall 2015	Spring 2016
0	42	57	16
1	32	49	17
2	73	94	15
3	126	200	71
4	73	171	53
5	23	47	17
6	2	12	9
Mean	**2.628**	**2.902**	**2.9471**
Standard Deviation	**1.386**	**1.418**	**1.3595**

ingly, there was not a strong linear correlation between the multi-step problem grade and the number of clicks in the CITA tutorials in any of the semesters. However, there was a significant difference between student's reported use of CITA. Table 4.10 shows the results of the chi-squared tests comparing students with different answers on question one of the exit survey.

Table 4.10
Scores on the multi-step problem for each semester were divided into groups by responses on question one of the survey. A chi-squared test was used to determine if the difference in the distribution of reported uses was statistically significant.

Semester	Chi Squared	P-Value	Cramer's V
Fall 2015	45.6616	0.00486	0.14760
Spring 2016	41.6404	0.01415	0.24530

The spring semester of 2016 shows a large difference in scores between students who report using the CITA tutorials and those who do not. Those who report using the tutorials more tend to perform better on the problem.

4.7.3 Student Opinions

The final part of the analysis of the CITA system involves a mixed-method study of student opinions about the system as well as how they used it during the semesters. First, we present the statistically significant results of the remaining questions on the exit survey. Then, we elaborate on the accompanying qualitative analysis of the comments and focus group sessions.

Exit Survey

Overall, student feedback on the exit survey was mostly positive. With each passing semester, students tended to use the CITA system more and more (question one). This was largely due to the fact that the system became more focused and sophisticated as time went on. Additionally, all of the semesters generally liked the overall system (question nine). The spring 2016 version of the system was more positively received than the summer 2015 version, which was more positively received than the fall 2015 version.

As we might have guessed from the analysis above, students in the on-campus section of the fall 2015 semester commented that they used the interactive walkthroughs more than the online sections. However, there was no statistically significant difference in the reported use of the tutorials in the summer 2015 or spring 2016 semesters. It is worth noting that this is a seemingly contradictory result when compared with the results from above. However, this result might be due to students over-estimating how much they use the system as well as low number statistics in the online sections. Figure 4.15 gives a follow-up comparison of the maximum number of nodes visited by a student in a CITA tutorial in the spring 2016 semester. A node is defined as a location in a homework problem where students click the "Enter" button; for example, this can occur after they submit an answer to the main question, after they complete a step in the tutorial, or after they read through a passage of text. We see that online students visit far fewer nodes than their on-campus colleagues.

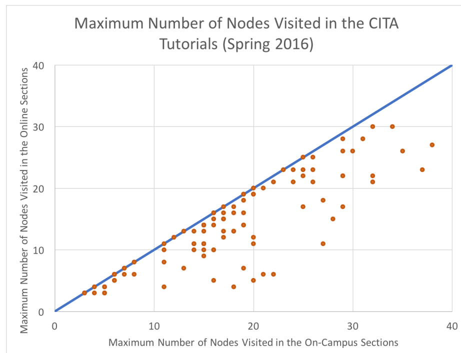

Figure 4.15. A comparison of the maximum number of nodes taken by a student within a CITA tutorial on a question-by-question basis. Each point in the graph represents a single question on the homework, while the x- and y-axes represent the maximum number of nodes visited by a student in the online and on-campus sections of PHYS 24100. The blue line in the diagram marks the spot where online and on-campus sections use the tutorials equally.

Students in PHYS 24100 and 24100D are allowed to work together on the homework. We provide Piazza classrooms for students to use should they want to ask questions but are not able to make our office hours. Overall, when students were asked if the online tutorial system improved communication with their colleagues, students were mostly neutral or slightly disagreed. There was an interesting division between gender and nationality. Female students generally disagreed more when asked if communication improved while males were more neutral. However, international students, on average, answered that the tutorials increased communication between them and their colleagues.

Generally speaking, students did not use the followup questions that were included at the end of the problem. Although many students commented that the followup questions helped them learn the material, they rarely used them afterwards. However, foreign students tended to complete the followup questions more than their on-campus colleagues. Thus, foreign students were far less likely to treat the followup questions as "optional" in all three semesters.

Qualitative Results

Our qualitative analysis complements our quantitative results very well. Overall, students were very happy with the work done on the CITA system. Most of the comments in the exit survey were positive in the semesters when it was administered (summer 2015, fall 2015, and spring 2016). In fact, we noticed that students used their previous online homework experience as a baseline to judge CITA on CHIP. One student commented that he likes the CITA tutorials better than other online homework systems that he has used in the past:

> The online homework system we used seems to be better than many of the other systems I have tried throughout my education. Not only is the error tolerance for answers great, but the homework system tries to help you through each problem. (Larry E., Summer 2015)

This was not an isolated case. Many students commented that they like the CITA format of tutorials more than other online systems and found the content generally helpful. We actually had a number of students comment that they liked CITA on CHIP better than WebAssign (the system that was used in PHYS 17200):

> The interactive walkthroughs are a very useful tool and why I like CHIP better than WebAssign. (Burt H., Spring 2016)

Furthermore, students constantly encouraged us to create tutorials for questions that were not covered under the current version of CHIP in their respective semester. Since a small group of researchers and professors was developing the tutorials, we could not create 138 new tutorials for 138 problems in every semester due to the time constraint. Additionally, it was not appropriate to design tutorials for some of the conceptual questions (e.g. true/false questions). Thus, we chose the iterative design process to provide constant improvements to the system as described earlier in the paper.

Although these may have been pragmatic reasons, this caused some cognitive dissonance for some. Many students lamented the fact that some of the questions did not have tutorials in the summer and fall semesters of 2015. We found from focus group interviews that during these times, students often fell back on online sources like Google and Yahoo Answers. It appears that if we want to include questions in the CITA system that do not include any sort of tutorial, we need to take special precautions to prevent cheating; for example, we might need to include a "primer" to prepare students for the analysis or change the phrasing of the question in each semester.

Although the majority of the class liked the CITA system, our tutorials were not always positively received. Some students felt like the focused feedback was not aimed at their skill level. This student in the fall 2015 semester commented that certain tutorials did not go deep enough into the material; these tutorials were updated for the spring 2016 semester.

> I felt very frustrated a lot of the times working with the online homework, just because, I always felt like there was a lack of feedback that it gave me. I would always ask help from my peers on doing the homework, but most of them were struggling just as much as me. (Carl A., Fall 2015)

The scaffolding that was provided in the fall 2015 semester was not sufficient for this particular student and many others. Due to the fact that we did not accurately predict his skill level and scaffold accordingly, the student did not find our system particularly helpful.

The number of negative comments about the system seems to decrease from semester to semester. This leads us to believe that we are becoming more effective at identifying student errors and filtering responses in order to guide students to the correct style of scaffolding. The spring semester of 2016, with its additional level of filtering, was particularly effective at judging student skill levels and providing feedback accordingly in the campus sections. However, the system is by no means perfect. More work is needed to identify and aid students who either have not attended lecture or who did not understand the lesson.

Another major strengths of CHIP was its cost. We were actually very surprised at how many students commented that they liked the system, not only because it was useful, but also because it did not cost them anything extra. Although it is unreasonable for a company to provide a service for free, it is important to note that college students do have a heavy financial burden placed upon them. Many students feel like it is not worth it to buy a book or a subscription to a system which will be used for four months and then abandoned.

Students repeatedly provided two suggestions for improving the CITA system. The first is to improve the overall coherence of the PHYS 24100 course. Even though we are pulling CITA examples from the notes and textbook, there still seems to be a huge gap between learning the concepts and applying them in actual problems. One student below sums this up in his comments on the course as a whole:

> Honestly, this course was terribly frustrating, as the majority of the homework questions did not line up with material taught in lecture, as the lectures were either entirely conceptual, or the examples were only for one very specific situation, neither of which was in the homework. (Andres M., Fall 2015)

Another student is slightly more positive about the situation, but still finds lectures to be rather unhelpful when compared with the textbook and tutorials.

> The homework system used in this course was very helpful, especially the interactive examples. However, lecture just made me very confused and I rarely learned anything from lecture. I learned the most from reading the

> book and the interactive examples on the homework. (Sean B., Spring 2016)

Many students found that the questions on the homework were much more difficult than the questions on the exam, not necessarily because the solution was more difficult, but rather because the homework questions asked for more detailed derivations while the exam questions asked for more general, conceptual understanding of the material. This suggests that a concept transfer study might yield very interesting results in future iterations of the system.

Moreover, as the CITA system became more sophisticated, students demanded that the remaining components of the course follow suit. Although there were some steps that we could take to cohere the homework and the other parts of the course (e.g. cite specific course notes, refer students to specific passages of the textbook, etc.), we had to be careful not to confuse the teaching of the material with scaffolding within a problem. Students were expected to have learned the material before attempting the homework. Although specific examples are useful to gain a better understanding of a concept, one must understand both the theory and application.

Finally, in the exit survey and focus group sessions, students often commented about time constraints. Engineering students have a difficult course load, and physics is just one of many courses taken. Many students, particularly in the fall 2015 semester, commented that the interactive walkthroughs were very extensive (positive), but very time consuming (negative). Some students commented that it was frustrating to go through a lengthy tutorial, only to find that a simple assumption made earlier in the problem was the culprit. One of the common suggestions that students offered to fix this problem was to decrease the length of the tutorial - skip the "worthless" setup of the problem and get to the relevant equations.

We did not take this approach when trying to tackle the problem of length. Instead, in the spring of 2016, we added the filter, asking students where they were in the solution of the problem. Based on their answer, they would either get a long, detailed tutorial or a short, general tutorial. Although the comments about time and efficiency did not disappear completely, we did see positive changes:

> The interactive homework system was extremely useful in both teaching the material and helping me finish the homework in a timely and efficient manner. (Collin O., Spring 2016)

4.8 Discussion

The results indicate that a variety of students used the CITA tutorials. This included different genders, nationalities, and previous skill-levels in physics. Thus, CITA use was not restricted to any one individual group. We did notice that on-campus students tended to use the CITA tutorials more than their online counterparts. Further analysis is needed to determine why this difference exists between the sections.

The results also indicate that CITA improves student performance in PHYS 24100 and PHYS 24100D. Students who navigated the CITA tutorials tended to have higher course grades, exam scores, and multi-step problem scores than similar students who did so less frequently. These results varied between a small and medium effect size. However, the improvement was not a simple linear relationship between clicks and gain; rather, the increase in scores leveled out after a certain number of clicks.

Our study suggests that students are motivated by the desire to efficiently solve the homework rather than learn physics problem solving skills. In fact, some students commented on an often used (but very ineffective) learning strategy - complete the homework as fast as possible and then learn the material when studying for the exam. Students also comment that online sources like Google and Yahoo Answers are common resources while they are doing the homework.

We certainly cannot claim to have solved this problem. Students are still motivated by good grades and completing the assignment. However, we did find that the focused feedback offered by Shallow CITA, Immersive CITA, and Postscripts improved student performance on a number of assessments. The first version of CITA seemed to do very little for students. However, versions two and three did show statistically significant growth in learning and problem solving.

Prior to the development of the CITA system, we noticed that online students tended to under-perform their on-campus colleagues (in some cases by almost half a letter grade). We found that those who used the system improved against those who did not. However, online sections tended to use the system much less than their on-campus counterparts - even though they might benefit from the tutorials more.

Figure 4.15 is a particularly vivid illustration of this fact. In the spring semester of 2016, online students were traversing only half or a third of the number of tutorial nodes that their on-campus counterparts traversed. This suggests that they either used the "fast track" more than the on-campus sections or they exited tutorials early. Either way, work is needed to tailor the tutorials to fit the online classes more effectively. This might involve conducting additional focus group sessions to determine how students specifically in the online sections use the tutorials.

We found that the followup questions (i.e. the Postscripts) were extremely underused in the CITA system. Postscripts outlined the final step of Polya's problem solving framework and reviewed the major concepts that were used in the problem. However, since students were trying to finish the homework assignments as quickly as possible, the vast majority of the class skipped this final moment of reflection. Future versions of CITA need to enforce this step of the problem solving process. We might consider making the Postscript worth some small number of points to encourage students participation.

It should be noted that the followup questions were answered more often by international students than by domestic students, and the international students commented that the tutorials increased communication between them and their classmates. We speculate that successful international students may have a stronger academic work ethic that their domestic counterparts. Thus, they might be more inclined to complete optional parts of the problem, even though they are not for credit. In addition, English may be a second language to many international students; thus, they are more inclined to read the entire text offered by the tutorials and follow-up questions to ensure that they understand all of the technical language associated with the problem.

Finally, we noticed that the actual number of clicks within a CITA tutorial might not be the best indicator of student learning. Aggregate results over an entire class never yielded any sort of useful result. Rather, we had to break the class up by some parameter, be it gender or previous grade, to yield some useful data. Even then, the actual number of clicks within a tutorial did not always yield a clear relationship. Sometimes a student's reported number of clicks was more indicative of success than their actual usage.

Since our tutorials were voluntary, students had different ideas of what heavy and light use meant. In fact, we saw that some students clicked within the system hundreds of times over a semester while others clicked thousands of times (a full order of magnitude increase). However, in the case of the multi-step problem, the reported use of CITA yielded far clearer results than the actual number of clicks. While most of the time the recorded number of clicks and the reported use of CITA correlate reasonable well, we recommend that researchers trying to probe optional questions in a tutorial system collect data on both the actual usage and students' perception of their usage.

4.9 Conclusions

We have built a set of interactive tutorials to guide students through their homework in PHYS 24100 (the introductory electromagnetism course for engineering students at Purdue University). By providing three different types of scaffolding - Shallow CITA, Immersive CITA, and Postscripts, we are able to filter student input and provide focused feedback in a more efficient way than any of the three scaffoldings acting alone.

Overall, the first version of CITA yielded little to no gains in student performance in the class. This is most likely due to the combination of the rapid pace of the summer class coupled with the fact that only half of the homework problems had any tutorial. However, we saw that there are small but consistent gains for students that use the CITA system in fall 2015 and spring 2016 versus those students that do not. This is especially true in the spring semester of 2016; certain exam questions showed improvements from the fall and there is noticeable improvement on a diagnostic multi-step problem when compared with the spring 2015 and fall 2015 semesters.

Due to the fact that students could choose whether or not to follow a tutorial, we found that a simple count of the number of clicks could not entirely measure student engagement. Rather, by combining data about the number of clicks taken by the student as well as his or her perception about how much he or she used the system, one could paint a much richer picture about the tutorials.

Work on an online homework system is never truly complete. There are always improvements that can be made to fit the changing requirements of the administration, teaching staff, and student body. We believe that we have built a solid platform upon which future iterations of the CITA system can be designed. Thus, there are many future directions toward which this project can progress:

1. **Increase Number Statistics:** We have only collected data from one year of PHYS 24100, and two semesters of CITA version 3. It would be useful to collect a larger sample of students using CITA version 3 so that we may get a better understanding of its efficacy.

2. **Implement "Dead End" Style Tutorials:** True problem solving is not a linear, step-by-step process. It would be useful to guide students through mistakes and "dead-ends" in the analysis in a controlled way so that they could learn from their mistakes rather than fear them.

3. **Teach Math Strategies:** Most of the problems that are addressed in a calculus-based introductory electromagnetism course can be reduced down to

simple algebra. However, many students report trouble with the calculus concepts presented in the course. Students' belief that their deficiency in calculus is preventing them to solve certain E&M problems may be able to be dispelled by CITA, which takes them step-by-step in filling the gap between calculus and simple algebra that is needed to solve the problem.

4. **Analyze the Graphical Design of the Tutorials:** We must contend not only with the content of a tutorial but its presentation. This is especially true when tutorials contain multiple branches of analysis. Which style of the graphical user interface yields the greatest increases in student learning?

5. **Bridge the Gap Between Introductory Physics Courses:** Electromagnetism is, at its heart, a field theory. Field theories can be extremely difficult for introductory students to grasp due to their abstract nature. By providing a connection between more familiar concepts (e.g. introductory mechanics) and new topics (e.g. electromagnetic fields), we may be able to make students more comfortable with the content of PHYS 24100.

4.10 Acknowledgments

This project is supported in part by a grant from the College of Science at Purdue University. We would also like to thank the PER@Purdue research group for their helpful discussions and insights.

4.11 Appendices from the Paper

4.12 Multi-Step Problem and Rubric

4.12.1 The Multi-Step Problem

In order to judge the efficacy of the interactive tutorials in CITA, we needed to analyze how students work through a non-trivial problem that was not directly connected to any of the homework assignments. Thus, we used a multi-step problem in a similar manner to the analysis done by Bruce Sherwood [103].

The following problem was given to on-campus students in their recitation class during the week in which they had just learned about power, intensity, and radiation pressure in lecture. At the beginning of recitation the teaching assistant gave a 20 minute review of these major concepts. Then, he or she had the students complete the following problem under test-like conditions (i.e. no book, no notes, no collaboration).

> A laser beam of power P = 10.0 W and diameter D = 1.00 mm is directed upward onto one circular face of a perfectly reflecting cylinder. The cylinder levitates due to the balance between the upward radiation force and the downward gravitational force.
>
> If the density of the cylinder is 1.25 g/cm^3, and the diameter of the circular face is 0.50 mm, what is the height H of the cylinder?

There were three main reasons for choosing this particular problem. First, it is a multiple step problem that combines electromagnetism concepts with free-body diagrams and density. Thus, it allows us to see if students can pull together different concepts (some well known, others just recently learned) and apply them in their analysis. Second, the problem had to be administered during a recitation that fit the class schedule. The recitation during which this problem was administered was conveniently scheduled right after students learned the needed concepts. Finally, it contained information just on the boundary of their knowledge. We did not expect any of the students too be completely comfortable with concepts like power, intensity, and radiation pressure. Rather, we wanted to test if students could reason through a problem, even if they did not entirely understand the domain of the analysis.

4.12.2 The Rubric

In order to grade the multi-step problems in a consistent way, we developed a standardized rubric. The solution was divided up into three different parts, each worth up to two points:

1. Setting up the statics problem using Newton's laws (or conservation of momentum)
2. Calculating the radiation force using the relationships between power, intensity, and pressure
3. Calculating the height of the cylinder using the definition of density and the geometry of a cylinder

Additionally, we made a note of any students that did not start off their solution with a diagram and any students that tried to calculate the time rate of change of momentum ($F = dp/dt$) since these were techniques taught to them in their previous physics class.

The development and testing of the rubric was an iterative process spanning two weeks. Once a draft of the rubric had been created, it was tested on a sample of about 30 solutions from the spring 2015 semester. The rubric was then adjusted in order to calibrate the scores to a mean between three and four. All in all, the team reviewed about half a dozen iterations of the rubric before deciding on its final form.

The detailed rubric is shown in Table 4.11. All students start with a score of zero. Then, the grader moves down the rubric row by row, determining to the best of his or her ability if the student has shown the work required to earn a point. If the grader determines that the student has shown the required work, he or she will increment the student's score by one and move onto the next checkpoint in the next row.

The student must perform a step completely correct in order to receive credit. For example, if a student writes out the correct formula relating density and mass, but does not get the calculation correct due to a unit error, he or she will not receive credit for "Solve for Mass". However, carry-over work is given credit. For example, the simple math error from above means that the student did not correctly "Solve for Mass". However, the student might still correctly "Solve for Height" using the mass that he or she calculated from earlier.

Value	**Setting Up Statics Problem**
0	Completely Incorrect or No Attempt
+1	Identify Gravitational Force on Cylinder (F = mg)
+1	Set Up Statics Problem or Draw Free-Body Diagram
Value	**Solving for the Radiation Force**
0	Completely Incorrect or No Attempt
+1	Solve for Intensity Using Power
+1	Solve for Radiation Force Using Intensity
Value	**Solving for the Height**
0	Completely Incorrect or No Attempt
+1	Solve for Mass Using Density
+1	Solve for Height Using Mass

Table 4.11
Rubric for the analysis of the multi-step problem. In addition to these six possible points, graders made a note of whether a student drew a diagram within their solution and whether a student tried to apply the time rate of change of momentum in the solution (F = dp/dt).

The solutions from each semester were pooled and graded together in order to minimize any drift in grades or biases from the grader. Additionally, the grading was spread out over a few weeks in order to prevent fatigue. A second grader reviewed a small sample of the problems and compared them to the original grader to ensure consistency. After all of the problems had been graded over all of the semesters, the team went back and compared the final scores of students with the same answer. If there was a grade discrepancy of more than one point, we reviewed the work and made a case-by-case decision on an appropriate final score.

4.13 End-of-Semester Survey

During the final week of the semester before final exams, students were asked to fill out a survey about their experience with the CHIP homework system. The nine questions of the survey asked about four major topics (we used factor analysis to confirm the groupings of the questions). Questions 1, 2, and 9 probe student's opinions of the utility of the CITA tutorials. Questions 3 and 7 probe students opinions about their workload during the semester. Questions 4 and 8 probe students ideas about communication with their colleagues during the semester. Finally, questions 5 and 6 probe student's opinions about the utility of the postscripts and followup questions.

Questions 1 - 8 could be answered on a five point Likert scale ("Strongly Agree" to "Strongly Disagree") with the additional option not to disclose any information ("I Prefer Not to Disclose"). Question 9 asked for student's overall satisfaction with the system, so the options instead became "Very Satisfied" to "Very Dissatisfied" plus the option not to disclose any information. Question 10 allowed students to write out any comments that they might have on the system.

The full survey along with the instructions is shown below.

We are constantly trying to improve the online homework system and would like your feedback. Please assess each statement using the drop down menu. Your choices are "Strongly Agree", "Agree", "Neutral", "Disagree", and "Strongly Disagree". If you would prefer not to answer a question, please mark the choice labeled "I Prefer Not to Disclose".

Please be sure to make an entry for each question before clicking the submit button. Once you submit a choice for a question, your answer cannot be changed for that question. However, if you submit your choices for only some of the questions, you can come back to this survey prior to the deadline and fill out the remainder of the choices later.

Interactive Walkthroughs (i.e. The Help Buttons)

1. I used the interactive walkthroughs to help me with homework problems.
2. The interactive walkthroughs helped me learn the material.
3. The interactive walkthroughs increased my workload for the semester.
4. The interactive walkthroughs increased communication between me and other students.

Followup Questions (i.e. The Purple Box Following Each Question)

1. I tried to answer the followup questions on each homework assignment.
2. The followup questions helped me learn the material.
3. The followup questions increased my workload for the semester.
4. The followup questions increased communication between me and other students.

General Questions

1. What is your overall satisfaction with the interactive homework system?
2. Do you have any general comments that you would like to share? Would you be willing to talk to us to provide more details (if so, write down your name and contact information)?

5. COMPARISON OF SCAFFOLDINGS

5.1 Preface

In the previous chapter, I provided an in-depth look at the CITA system and how it influenced student performance in PHYS 24100. I found that the system did improve student performance in the physics class, with the more sophisticated systems tending to yield better results. Additionally, I found that there was a generally positive opinion of the CITA system in all three semesters where the exit survey and focus group sessions were given. However, we must dive deeper. Why are the fall 2015 and spring 2016 results better than summer 2015 and before? How do students actually traverse the individual nodes of a CITA tutorial and learn physics concepts?

This chapter is also a paper destined for publication. Again, at the time of the publication of this thesis, it was in pre-print format. We conducted an analysis of the nodes of the CITA system and determined patterns that students exhibited when using the tutorials.

5.2 Introduction

An instructor's job does not stop after a class session has ended. Educators need to provide students with additional assistance outside of class as they work through homework assignments. Oftentimes, helping individual students requires a large amount of time and energy - especially in large universities where an introductory physics class can have an attendance in the hundreds or even thousands. Thus, many instructors turn to online homework systems to administer assignments, grade problems, and provide supplementary instruction. The requirements for these online systems are constantly increasing; these systems are expected to identify student errors, provide useful feedback, and sometimes even guide students through the analysis.

Over the past two years, the research group (henceforth known as "we") has developed a Computerized Interactive Teaching Assistant (CITA) for the calculus-based, introductory electricity and magnetism (E&M) course for engineering students at Purdue University. The goal of this system is to guide students through homework problems while reinforcing problem solving techniques. CITA, like many other tutorial systems, takes students through a series of manageable steps of analysis in order to demonstrate a problem solving process. Each tutorial is completely optional - students can choose to traverse a tutorial if they are having trouble with the problem at hand. Many studies have been conducted on how different structures and styles of scaffolding affect student performance in physics, both in an online and live environment [121–126]. These studies demonstrate how different types of scaffolding can be effective in different contexts.

One of the defining features of the Internet is the easy access to information. One can navigate between different sources with a click of a button. Thus, we must remember that the educational scaffolding that we design is not used in a vacuum. Instead, it competes with other sources for student's attention. We seek to explain how students navigate our designed scaffolds in an online environment where a plethora of other options are available. Our research questions are:

1. What are the patterns of use of the CITA tutorials? Do students have a preference for one type of CITA structure over another? If so, why is it preferred?

2. Are there specific points in the CITA tutorials that cause students to turn to other sources? If so, how can these be avoided?

5.3 Background

5.3.1 Description of the Class

CITA was developed for the calculus-based, introductory E&M courses at Purdue University, entitled "Electricity and Optics" and coded as PHYS 24100 and PHYS 24100D. These courses are meant for sophomore engineering majors, with the exception of electrical engineering majors who take a different physics course than the rest of their colleagues [9].

The "D" in PHYS 24100D is used during the fall and spring semesters to signify a distance learning (i.e. online) section of the course. However, the online and on-campus sections of Electricity and Optics cover the same material: electric charge, electromagnetic fields, Maxwell's equations, geometric optics, and interference effects. All students use *Physics for Scientists and Engineers* by Tipler and Mosca as their main textbook [10]. The sections differ only in the fact that the online classes watch video lectures and attend online recitations through Cisco WebEx [11] while the regular sections watch live lectures and attend live recitations on campus.

Engineering students usually take PHYS 24100/24100D in the fall semester (also known as the "on-semester"). Alternatively, engineering students can choose to take the course in the spring or summer if they are ahead or behind in their schedules (also known as the "off-semesters"). The option of taking Electricity and Optics in the summer has grown in popularity ever since the online version of the course was designed. Table 5.1 shows the total enrollment in the campus and online sections of Electricity and Optics over the last three years. During each of the fall and spring semesters, just under 10% of the total class is enrolled in PHYS 24100D.

Electricity and Optics is generally taken after students complete PHYS 17200 - Modern Mechanics. The introductory mechanics course is taught using the first semester of the *Matter and Interactions* curriculum by Ruth Chabay and Bruce Sherwood [12].

Table 5.1
Enrollment in the campus and online sections of Electricity and Optics from spring 2014 through summer 2016.

Semester	Campus	Online	Percent Online
Spring 2014	465	41	8.1%
Summer 2014	0	153	100.0%
Fall 2014	725	62	7.9%
Spring 2015	474	43	8.3%
Summer 2015	0	187	100.0%
Fall 2015	728	69	8.7%
Spring 2016	339	36	9.6%
Summer 2016	0	222	100.0%

5.3.2 Scaffolding and Problem Solving

The learning theory that is the basis for our design and development of CITA is constructivism. This theory asserts that when a student learns a new concept, he or she always has some prior knowledge related to the subject. This knowledge can come from formal study (e.g. taking a science class in high school) or from out-of-the-classroom experiences (e.g. learning a painful lesson about gravity after falling to the ground). It includes normative models of how the world works as well as non-normative models of the universe (e.g. gravity always points "downwards"). No matter if these assumptions are correct or incorrect, they stem from the fact that students notice regularities in the world that are confirmed by repeated experience [21, 22].

Constructivists believe that learning is an active process of building and revising models, where new information about the world is assimilated and/or accommodated into previous models about how the world works. Thus, teaching strategies must facilitate students' development of a fundamental understanding of the material; simple memorization is usually not enough to change one's models. Within the realm of physics education, the goal is not only for students to generate normative models of the universe, but also to apply these models to solve real-world problems.

Most introductory undergraduate students can be classified as "novice" problem solvers. They do not have a well-defined methodology for approaching a given problem. Instead, they employ a variety of "epistemic games" to connect seemingly unrelated concepts [43]. These games might include working backwards from a known solution without identifying the fundamental laws governing the analysis and identifying similar features between problems in order to look up a solution [41, 42].

Conversely, master problem solvers have a well-defined strategy to solve any given problem. They start from first principles and work their way towards an answer in a structured way [45]. They follow the basic steps of defining the problem, devising a solution, working towards the answer, and checking the output, as described by Polya [39]. Experts have a thorough understanding of the laws of physics, and thus feel comfortable starting from first principles. Novices do not necessarily have a completely accurate model of the universe, so they need learning experiences through which they can relate new knowledge to existing knowledge. Instructors may provide instructional scaffolding to help students develop new models and revise existing models in the learning process.

Instructional scaffolding provides a way for students to build an understanding of concepts that may initially be beyond their skill level [48]. Lev Vygotsky initially proposed that children can accomplish tasks outside of their current skill level with the assistance of an adult. The adult creates a structure of small steps that the child can easily follow. Then, the child connects each step to arrive at the final answer. This is applicable to both physical tasks (e.g. learning to ride a bicycle) and mental tasks (e.g. learning about right and wrong through reward and punishment). Vygotsky called the area between what is known and what is not known the Zone of Proximal Development (ZPD). Learning takes place in the ZPD - otherwise the task at hand is either too difficult or is already known by the child [30, 117].

Education research has adopted the idea of the ZPD and applied it to students' learning. Scaffolding is an umbrella term that refers to any technique that is used to move a student from their current understanding of the world towards an understanding of new material. As referenced in the introduction, scaffolding comes in a huge variety of forms and styles. There is no one "correct" scaffolding for all students because learning is an inherently individual, social, and subjective process [48]. However, all scaffoldings break a complex problem down into manageable pieces in order to help students traverse the ZPD. Appropriate scaffolding in physics problem solving, for example, provides students with a solid conceptual background of the problem that allows them to relate sophisticated mathematical concepts to observations in nature [51, 52].

5.3.3 Scaffolding in CITA

CITA is built upon Purdue's in-house homework system called CHIP (Computerized Homework in Physics). CHIP is based on the online homework system called CPlite that was originally developed at the University of Illinois at Urbana-Champaign (UIUC). CITA is designed to be a sequence of successively more detailed tutorials that target students with a variety of skill levels. Our early qualitative analysis of participants in the study indicated that students desire focused help right when they encounter a problem. Otherwise, they will turn to external sources like Google or Yahoo! Answers (which will obviously detract from the learning process). Thus, the tutorial system is based on a pair of key ideas:

1. Provide students with focused feedback to any incorrect answers.

2. Give students the power to easily find and navigate a tutorial that fits their skill level.

In order to build this scaffolded system, we divided up the structure of CITA into three main parts: Shallow CITA, Immersive CITA, and Postscripts. These parts are meant to work together in harmony. They take a student through Polya's problem solving process in a controlled way.

An example of Shallow CITA is shown in Figure 5.1. A student must solve a homework problem for a grade. If the student makes an easily identifiable mistake on the problem (e.g. a unit error, a sign error, a simple conceptual error) a box will appear and offer focused feedback on how he or she can correct the mistake to proceed. Shallow CITA is used to guide students who are close to a correct answer; thus, it is assumed that a student generally understands the major concepts within the problem when Shallow CITA provides feedback. If Shallow CITA cannot identify the mistake made by the student or if the mistake is rather complicated, it suggests that the student attempt an Immersive CITA tutorial.

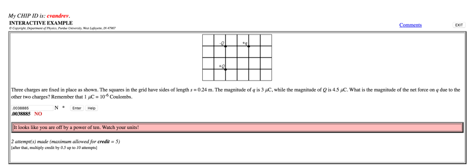

Figure 5.1. A screen shot of a homework problem in the CHIP homework system demonstrating Shallow CITA feedback. The red box appears after a student enters an incorrect answer that matches a common error.

Immersive CITA provides problem-specific, detailed tutorials (see Figure 5.2). During the fall of 2015 and spring of 2016, we tested three different structures of Immersive CITA.

The first structure was a step-by-step tutorial. Each step of the tutorial contained a few sentences of explanation followed by a question. Depending on the complexity of the problem, a tutorial contained between six and thirty steps (i.e. going through each step of the setup and mathematical analysis in a Gauss's law problem). Thus, students were not simply reading a transcript of the answer. Rather, they had to work towards the solution using the designed educational scaffolding.

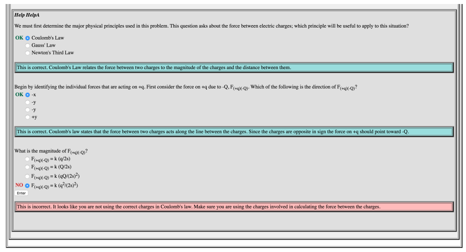

Figure 5.2. A screen shot of a homework problem in the CHIP homework system demonstrating an Immersive CITA tutorial. The tutorial guides students through the solution of the problem in a step by step fashion.

The second structure was a truncated step-by-step tutorial. Many of the problems on the homework repeated similar concepts or similar solutions. Thus, truncated step-by-step solutions simply outlined the major steps in the problem and/or referenced other Immersive CITA tutorials within the same homework assignment. Generally speaking, these tutorials contained between one and three steps, depending on the complexity of the problem. For example, a truncated step-by-step tutorial for a geometric optics problem might remind students to apply Snell's law at certain points on a prism and refer to a different Immersive CITA tutorial in the homework.

The final structure was the branching style. In this structure, the first step of the Immersive CITA tutorial asked students to rate their confidence with the problem and decide if they wanted a detailed, step-by-step tutorial or a faster, general overview of the problem. Depending on their answer, the system filtered the student body into one path or the other. The trade-off here is time and effort; one path is quick to view, but it does not offer all of the little details needed for the full derivation. The other path was very thorough and would often lead to a point where the answer was relatively simple to find. However, this path required much more time (sometimes fifteen minutes or more for some complicated problems).

Students could enter and exit the Immersive CITA tutorial at any time. Thus, the final step of Polya's problem solving process was not included in Immersive CITA because we wanted all students to go through it - whether they needed a tutorial or not. Figure 5.3 shows the final part of the CITA program - the Postscript. After a student correctly answered the homework question, the Poscript reviewed Polya's final step in the problem solving process. We asked students to think about the work that they just completed by offering simple variations of the problem. For example, if a student solved for the electric field at some point in space and time, the Postscript might ask what the electric field would be at some later time. During the spring of 2016, each of the postscripts included a "real world application" - an example where the setup, model, or solution of the problem can be found in the realm of engineering.

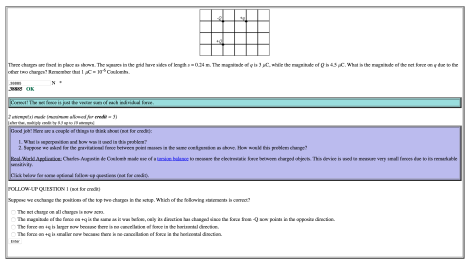

Figure 5.3. A screen shot of a homework problem in the CHIP homework system demonstrating a Postscript. The purple box at the end of the problem encourages students to think about their answer.

CITA was developed over the summer 2015, fall 2015, and spring 2016 semesters. Most of the problems in the homework system were not paired with any sort of tutorial prior to the summer semester of 2015; therefore, we consider the spring semester of 2015 a baseline semester. The summer 2015 semester consisted of about 50% no tutorial and 50% detailed step-by-step tutorials. The fall 2015 semester consisted of 40% detailed step-by-step tutorials, 40% truncated step-by-step tutorials, and 20% no tutorial. The spring 2016 semester consisted of just over 50% branching tutorials, 30% detailed step-by-step tutorials, and 20% no tutorial.

5.4 Research Design

This research project used a mixed methods design. Mixed methods research involves collecting, analyzing, and integrating both qualitative and quantitative data in order to gain a better understanding of a research domain than either approach alone. The mixing of qualitative and quantitative data can be very powerful for a number of reasons. First, the synthesis of data from different sources leads to greater confidence in the validity of the conclusions. Second, the answers to the research questions merge a number of perspectives, leading to a more complete conclusion with fewer "gaps" in the analysis. Finally, any pre-existing assumptions from the researchers are less likely to influence the results due to the fact that the conclusions must support both qualitative and quantitative data [94].

The study took place between the spring semester of 2015 and the summer semester of 2016 at Purdue University. It focused on the second semester, introductory physics courses for students pursuing an engineering degree - PHYS 24100 and PHYS 24100D. The "D" signifies a distance learning (i.e. online) section of the course.

5.5 Research Participants

Students had the option of following a CITA tutorial in order to learn how to solve a given homework problem. Alternatively, they could skip over a tutorial paired with a homework problem completely. Whether a student decided to use the CITA tutorials or not, he or she was asked to complete a demographics survey and exit survey that probed their usage of the system. At the end of the semester, students were given the option to elaborate more on their opinions of the system through voluntary focus group interviews.

Thus, students were only required to complete the surveys during the semester; CITA tutorial use and focus group sessions were completely voluntary.

5.6 Quantitative Methods

As a student works through the graded problem or an Immersive CITA tutorial in CHIP, his or her progress is automatically recorded by the system. Every time an answer is given or a passage of text is read, a student will click an "Enter" button in order to progress to the next step (or to finish the problem if the graded answer is correct). The CHIP system automatically records every student click with a time stamp and details of where it took place. This means that we can review the entire click history of a student as he or she works through a question in the homework. We collected the click history for students from the summer of 2015 through the spring of 2016 and compared how students navigate different structures of tutorials.

We leveraged a variety of quantitative techniques to uncover patterns in the data. Initial exploratory analysis included ANOVA and chi-squared tests to compare the click numbers between problems as well as visualizing the data with network digraphs. Matlab (version R2016a) was an excellent resource for this because it allowed us to visualize the data and quickly calculate graph statistics for each problem (see Figure 5.4).

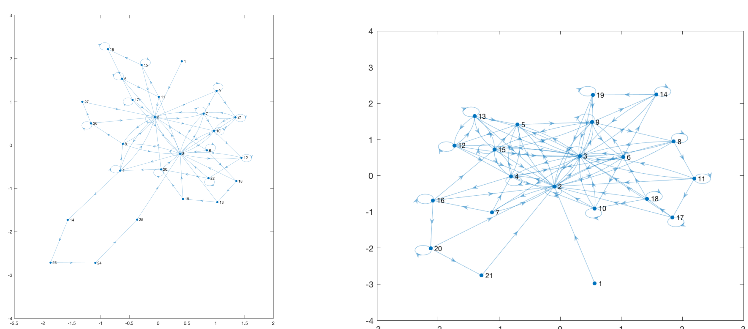

Figure 5.4. Examples of directed graphs generated using Matlab. Each of the points in the graphs refers to a specific step in an Immersive CITA tutorial or the graded question in the homework. The lines in the graph are weighted and represent the number of students that move from step to step in a problem.

After the exploratory quantitative analysis was complete, the results were compared with the qualitative data (described in detail below). Then, final conclusions were made.

5.7 Qualitative Methods

Strategy of Inquiry

The qualitative part of this study analyzed student opinions about the CITA program as well as their general study habits. The data came from three main sources. At the end of each semester, we asked students to fill out a survey about their use and satisfaction with the CITA system. The final question of the exit survey was an open space for students to make comments. Next, we used Piazza [109] over the semester as a forum to answer student questions. We could analyze the comments on the Piazza help system and see how they related to the tutorials. Finally, we offered students the opportunity to participate in focus group sessions at the end of the semester to share some of their opinions. The exit survey counted as a quiz grade (however, students had the option of answering “I Prefer Not to Disclose” on any question). Students were rewarded a ten dollar gift card if they completed a focus group session. Piazza offered no reward.

We used purposeful sampling as our design strategy for the focus group sessions. Patton defines purposeful sampling as one where the researcher chooses specific participants in order to gain insight about a specific phenomenon. Cases are selected because they are “information rich” [107]. When students signed up to volunteer for our focus group sessions, we asked them what their overall impression of the homework system was - “Generally Positive”, “Generally Negative”, or “Neutral”. We then divided the participants into groups based on their reported overall impression of the system and whether they were in an online or on-campus section. Thus, students were hopefully more comfortable sharing their opinions - good or bad - due to the fact that they were in a homogeneous group.

Our strategy for analyzing the transcripts from the three sources above consisted of a three-cycle coding plan. Our codes are based on Saldana’s reference [105]. First, each of the transcripts was analyzed separately using rounds of attribute coding and descriptive coding in order to get a sense of the recurring comments made by students.

Then, the transcripts were analyzed separately again using rounds of elaborative coding and pattern coding in order to synthesize the results of the first cycle and identify the major themes in the transcripts. The final cycle of coding pooled all of the sources together in order to determine patterns between semesters; it consisted of evaluation coding and longitudinal coding.

Roles of the Researchers

Researchers conducting a qualitative study must acknowledge that their presence can alter the answers given by the participants. Thus, they must take some time beforehand to determine how they might inadvertently skew the data as well as how to minimize this artifact [108]. In this study the principle investigators are Mr. Cyrus Vandrevala, Dr. Lynn Bryan, Dr. Andrew Hirsch, Dr. Hisao Nakanishi, and Dr. Laura Pyrak-Nolte.

Mr. Cyrus Vandrevala has been a teaching assistant for PHYS 24100 and PHYS 24100D since the fall semester of 2011. Additionally, he has coordinated the course during summer sessions and helped develop the online sections. Since he led all of the focus group sessions in the fall of 2015 and spring of 2016, he has not taken a teaching assistant position from the spring of 2015 onwards. Instead, he has a part-time assignment answering questions on Piazza because students will never meet him face-to-face.

Dr. Hisao Nakanishi is the “father” of CHIP at Purdue University. He is one of the primary developers of the system and has implemented the underlying structure upon which the branching Immersive CITA tutorials were built. Although he is not directly affiliated with PHYS 24100 or PHYS 24100D, he addresses student questions sent through the CHIP help system that pertain to errors in the problems, bugs in the codebase, or administration issues.

Dr. Laura Pyrak-Nolte has taught or coordinated PHYS 24100 and PHYS 24100D since 2005. As the course coordinator, she is in charge of creating exams, assigning final grades, and approving changes to the homework system. Additionally, she is the lecturer that is seen and heard in the online lecture videos shown to the PHYS 24100D class.

Drs. Lynn Bryan and Andrew Hirsch provided the educational theory that supports this study. Neither of them are directly involved with PHYS 24100 or PHYS 24100D, but Andrew Hirsch is one of the main instructors of PHYS 17200 - the introductory physics class that is a prerequisite for PHYS 24100 and PHYS 24100D.

5.8 Results

5.8.1 Overall Use of Immersive CITA

We first analyzed the use statistics of CITA tutorials. As mentioned above, the homework system started off with almost no tutorials (most of the problems had no extra scaffolding). Then, as time went by, we developed sophisticated structures that were paired with each problem. As described in the previous sections, different semesters used different structures of scaffolding.

When a student opens up an Immersive CITA tutorial, he or she does not know exactly what structure the tutorial will take. Additionally, the homework problems are the same between semesters. Thus, we would expect approximately the same fraction of students to enter the Immersive CITA tutorial in each semester for any given problem. We filtered the homework questions by semester and structure in order to determine if the fraction of students who enter an Immersive CITA tutorial changes between semesters. We used ANOVA analysis to compare the problems, and found no statistically significant differences in the means.

Additionally, we aggregated all of the problems and compared them on a semester-by-semester basis to see if there were any changes in usage. Table 5.2 gives the mean fraction of students that enter an Immersive CITA tutorial, averaged over all problems. ANOVA analysis once again showed no statistically significant difference in the means.

Table 5.2
Fraction of students that start an Immersive CITA tutorial, averaged over the semester and section.

Semester	Section	Mean	Standard Deviation
Summer 2015	Online	0.1387	0.1034
Fall 2015	Campus	0.2345	0.1170
	Online	0.1416	0.0900
Spring 2016	Campus	0.25543	0.1196
	Online	0.1057	0.0849

Figure 5.5 shows histograms of the on-campus data from Table 5.2. There is a wide range of values, meaning that some of the tutorials that were paired with problems were rarely, if ever, accessed while others were accessed by over half the class.

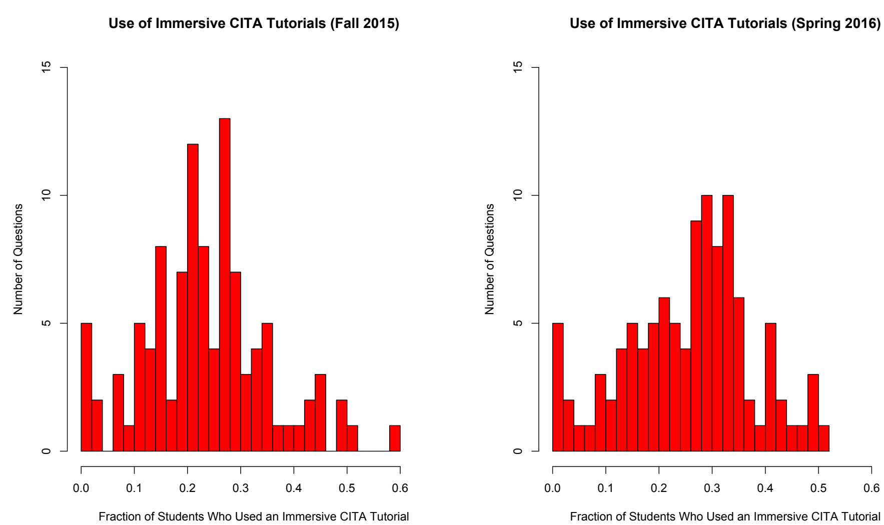

Figure 5.5. Histograms showing the use of CITA in the fall 2015 semester (left) and the spring 2016 semester (right). The horizontal axis shows the fraction of students who used an Immersive CITA tutorial while the vertical axis displays the number of problems.

5.8.2 Overall Use of Postscripts

During the focus group interviews, students repeatedly mentioned that they generally did not complete the postscripts. Specifically, they did not reflect on any of the scenarios given, and they did not complete any follow-up questions if the problem contained them. Their main reason for this was time constraints - students were most interested in quickly finishing the homework. Since the postscripts did not count for any points, they did not seem worthwhile.

We tested students' use of the postscripts in order to elaborate on this result. We collected 27 homework problems that contained follow-up questions in the postscript in the summer 2015, fall 2015, and spring 2016 semesters. These problems had identical or nearly identical text and follow-up questions in each of the semesters. Additionally, each of these problems was paired with a CITA tutorial - linear tutorials in the summer and fall of 2015, branching tutorials in the spring of 2016. In each question, we calculated the percent of students who attempted at least one of the follow-up questions.

Table 5.3
The fraction of students who attempted the first of the follow-up questions in the postscript, averaged over the 27 test problems.

Semester	**Section**	**Mean**	**Standard Deviation**
Summer 2015	Online	0.0999	0.0539
Fall 2015	Campus	0.1003	0.0600
	Online	0.0769	0.0624
Spring 2016	Campus	0.1649	0.0671
	Online	0.0867	0.0707

Table 5.3 shows the mean fraction of students who attempted at least the first follow-up question, divided by semester and section. The overall results mirror what students said in the focus group sessions - only a small percentage of the class com-

pleted the follow-up exercises. However, the on-campus section of the spring 2016 semester showed a 60% increase in the use of postscripts, averaged over the sample problems. ANOVA analysis of the on-campus sections shows a statistically significant increase in postscript usage with a p-value of 0.0005 and a large effect size (eta squared) of 0.2104.

The online students do not show the same growth in the use of the postscript. ANOVA analysis of the on-campus sections yield a p-value greater than the 0.05 cutoff. Thus, we must accept the null hypothesis that there is not change in postscript use between the online classes in different semesters.

5.8.3 The Truncated Step-by-Step Structure vs. the Detailed Step-by-Step Structure

All three sources of qualitative data show that students did not like the truncated step-by-step tutorials. Students commented that this style of scaffolding was not very helpful, and they turned to other sources. At best, this structure of tutorial was only helpful in simple problems that only required one or two steps of analysis. Students were especially annoyed if they felt like the tutorial was simply guiding them to a different resource with no extra content being offered by the help. In one group discussion where students were talking about simplifying the tutorials and just referencing the text, one student had this to say:

> I don't know... Like, the system in which you just remove the interactive tutorial and say, "Refer to the text". [...] I don't know, it's kind of demeaning in a way if you go to do that. Because it's like, oh I guess I just didn't read the textbook. [In reality,] I didn't understand the problem. (John M., Spring 2016)

When asked how this structure could be improved, a few students commented that it really was not worthwhile, and instead, all of the tutorials should have the detailed step-by-step structure. Some students commented that the truncated step-by-step structure could be more useful if we included a specific equation that related to the problem. Alternatively, students commented that it would be helpful to see the first few steps of the analysis (i.e. the start of a worked example) or a list of incorrect methods of analysis.

Conversely, comments were generally positive about the detailed step-by-step tutorials. Many commented that providing a few sentences of background and then asking a question at each step ensured that they understood the material before moving on. However, a couple of students commented that certain steps in the problem were very frustrating. One student elaborated on the check box style, "multiple answer" of step.

> I don't like where it's, like, "check which of these are true" when there's, like, five answers. It's, like, okay, I'm kind of in the help for a reason [and] if I just wouldn't get them right away, like, if I wouldn't get them after two or three attempts I'd just start trying to find what the answer is just so I can understand it. (Frank S., Fall 2015)

> I just found those [check box questions] frustrating personally, because like, okay, I'm in help for a reason. I'm trying to get help. (Samuel K., Fall 2015)

The student above commented that his strategy eventually became guess and check, just to get past that specific type of question. Quantitative analysis showed similar results. We found that numerical questions in an Immersive CITA tutorial were a point where students became frustrated. We will elaborate on some of these frustrations in the sections below.

Finally, a few students commented that the linear Immersive CITA tutorials could get very long; this was especially poignant if they had made a small error in the problem that was not caught by Shallow CITA and then had to go through a long derivation, only to realize that their mistake was trivial. However, many students were fine with the length of the tutorials. Going through the steps of the analysis made them feel more confident about their answer. These comments were the inspiration for the development of the branching structures.

5.8.4 The Branching Structure

We designed the branching structure of tutorial to try to give focused feedback to students who were using Immersive CITA. Overall, the branching structure was well received. Most students did not mind answering the filter question and proceeded with the tutorial. Students commented that the detailed branch was much more useful than the general branch, but some students used the general branch to check over their work if they felt like they were close to a solution. However, there were a couple of unexpected issues with the branching structure.

First the phrasing of the question that filtered students in to the two groups was important. During the development of the system, we tried to use an informal, second person style of writing (e.g. What can you do at this point? What is the final form of this expression after you have simplified it?). Thus, the filtering question asked students what their confidence level in the problem was and which structure of tutorial they would like to traverse.

One student commented that the personal tone was actually a deterrent for using the Immersive CITA tutorials at the beginning of the semester. When the tutorial asked her which path she would like to take, she thought that there was a right and wrong answer, and that she would be punished in some way for using the help.

> When I first saw that first question of, "How well do you know the problem?" The first thing that popped in my head was like, "Am I going to be penalized for saying, I have no idea?" (Sarah R., Spring 2016)

However, once she spoke with other students in the class and learned that no points would be subtracted from her score, she happily used the tutorials thereafter. Another student commented that the branch that gave a quick overview of the solution was not useful.

> I thought that [the branch which gave a quick overview] was kind of pointless for me. [...] It's just me but, if I know, I know. [...] I just didn't understand why would I, why would you, need this section. (Bill A., Spring 2016)

Many students who depended on the tutorials to step them through complicated problems had a similar opinion. If you are clicking on a help button, why would you need a short overview of the concepts? A follow-up quantitative analysis of this question shows some distinct differences between students in different sections. Figure 5.6 shows the fractions of the total student body that chose the general and detailed paths for each branching tutorial. The on-campus sections preferred the detailed paths over the general paths while the online students accessed the different paths more equally.

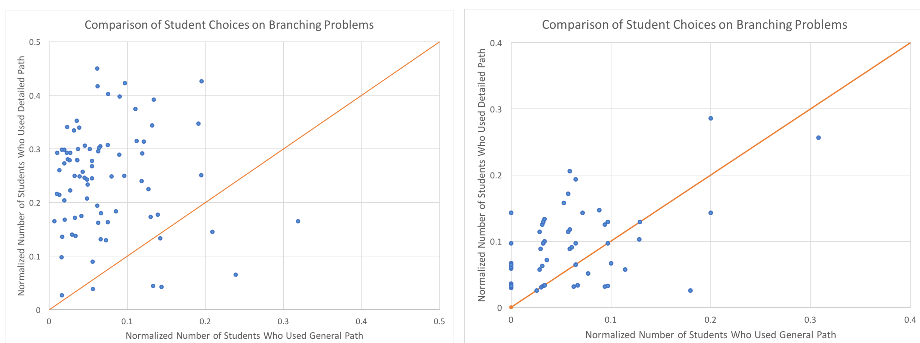

Figure 5.6. A comparison of the choices that students made when traversing the branching problems. The left graph shows the on-campus sections while the right graph shows the online sections. The number of students who traversed a given path in the branching problem was normalized with respect to the total number of students in the class.

A few patterns are evident in the graphs. Online students were accessing the shorter, overview paths much more often than their on-campus colleagues. In fact, the on-campus students preferred the more detailed paths than the overviews. Additionally, online students accessed the tutorials much less than their on-campus colleagues (whether they be general or detailed).

5.8.5 Student Retention as They Traverse a Tutorial

Students could freely enter and exit the Immersive CITA tutorials whenever they pleased. Thus, we were interested in finding out how far into a tutorial students navigated before leaving and attempting the graded homework question. In general, students reached the end of the truncated step-by-step tutorials because they were very short (only a few clicks at most). However, the detailed step-by-step tutorials and the branching tutorials yielded different results.

Overview Path of Branching Tutorial

Of the students who entered an Immersive CITA tutorial, we tracked how many of them proceeded to the next step. Then, we plotted the number of students at each step, normalized to the total number of students in the class. In the case of the branching structure, we did this individually for both paths. For the majority of the Immersive CITA tutorials, student retention rates fit a straight line with a negative slope. However, there is a sharp drop off of students near the end of the tutorial. For example, Figure 5.7 shows the fraction of students that completed the overview path of the branching tutorials through the final step (labeled General Path in the figures).

The data seemed rather disheartening as most problems showed almost no students reaching the end. However, we found that many students dropped out of the tutorial in the second to last step, due in part to the wording of the tutorials which indicated to students that they were coming to an end. When we repeat the analysis above with the second to last step of each tutorial, the graph changes significantly (see Figure 5.8). Students were very likely to work through the solution to almost the end.

The detailed step-by-step tutorials along with the detailed branch show slightly different results.

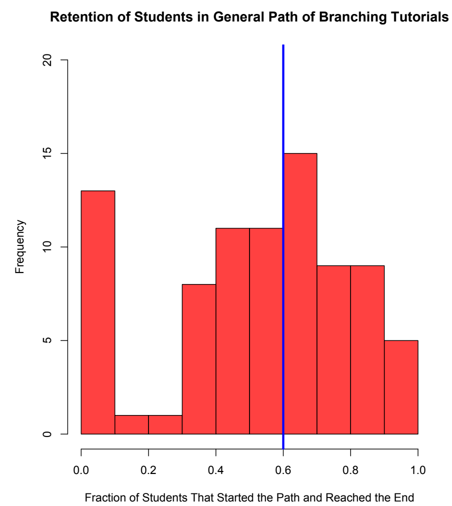

Figure 5.7. Histogram of the fraction of students who work the general path through to the end. The median is plotted as a blue line.

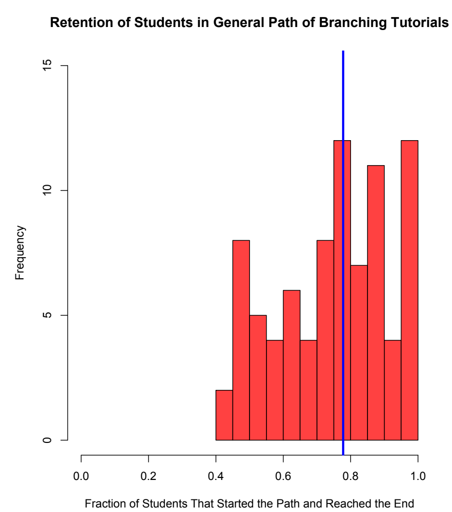

Figure 5.8. Histogram of the fraction of students who work the general path through to at least one step before the end of the tutorial. The median is plotted as a blue line.

Detailed Step-by-Step Tutorials and Detailed Branch

Our method of analysis for these structures was the same as that of the previous section - we tracked student clicks as they navigated the tutorials and determined how many students move onto the next step. Like the results in the previous section, we saw a similar pattern for most problems. Most of the tutorials showed the simple linear relationship between step number and number of students. However, a few of the problems contained a drastic decrease in retention from step to step. Figure 5.9 shows two examples of how the retention suddenly changed in two such problems.

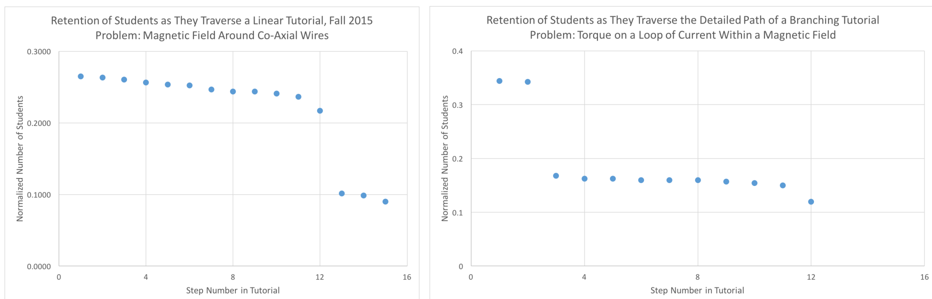

Figure 5.9. Fraction of students who completed each step in an Immersive CITA tutorial. There is a sudden drop in retention at certain steps of the tutorial.

Upon seeing these trends, we looked through the tutorials to identify why students were leaving so suddenly. We found that the steps where retention decreases severely correspond to specific types of questions - those that require a numerical answer, those that are a multiple-answer style question, etc. Student engagement in these tutorials are better fit with a series of step functions rather than a single linear function.

5.9 Discussion

Our research questions covered two major topics: (1) do students have a preference for one type of CITA structure over another? and (2) are there specific points in the tutorials that frustrate students and potentially cause them to exit prematurely? The students in PHYS 24100 reported that the detailed step-by-step and branching structures were preferred to the truncated step-by-step structure. The truncated structure did not provide enough support for students to comfortably traverse the problem. Additionally, we found that there are specific points in the CITA tutorials that caused students to leave prematurely.

Problems that were given the truncated step-by-step structure were widely hated by students. Many students commented that they did not address their specific issues or that they already understood all of the parts outlined in the tutorial but did not know how to apply them to the specific problem at hand. In these cases students turned to other sources like Google or Yahoo! Answers. Not surprisingly, students prefer tutorials that walk them through a problem in a step-by-step fashion, and thus, preferred the other structures.

However, a subtle balancing act took place in the linear and branching tutorials as well. Problems that were generally simple or straightforward did not merit the use of a tutorial. Thus, generally low numbers of students entered the tutorials on these problems. However, more difficult problems did merit at least a cursory glance at the tutorials. Once students were inside of the tutorial, their retention matched a downward step function (or a series of downward step functions). At first, large numbers of students moved through each step of the tutorial, and very few students left the lesson. At this point in time, the perceived gains from the tutorial outweighed the loss of time and expenditure of effort.

When a student reached a question that seems tedious (not necessarily difficult) another decision had to be made. Is it worth it to continue on through this step of the analysis or is it better to leave the tutorial and piece together a solution from what was previously discussed? Some students persevered and went through the tedium. However, a large number of students left the tutorial at this point. This is the sharp downward slope of the step function. Many students stated that they prefer to have tutorials bundled with each specific homework problem rather than having to search for solutions online.

We found that steps in the CITA tutorials that required numerical answers were largely responsible for students leaving the tutorial. This was especially true if the tutorial highlighted an equation before the numerical step - even if the equation was not the final formula needed for the problem. This effect can be explained as expectation violation. Students who use the CITA system expect a highly focused and efficient tutorial. However, once a tedious step occurs, the student's expectation seems to be violated, and he or she drops out of the problem. We saw no preference for certain lengths of tutorials. On the one hand, students worked through to the end of tutorials with upwards of 20 steps. On the other hand, students left a tutorial early when it had just a few steps, but one seemed tedious.

5.10 Conclusions

Our study of the different tutorial structures shows that students prefer the detailed step-by-step and branching structures to the truncated step-by-step structure. They commented that the truncated step-by-step tutorial failed to provide the necessary support needed to complete the problem. From an educator's standpoint, this structure did not effectively scaffold the material and techniques needed to solve the problem. Conversely, the detailed step-by-step and branching structures provided enough support for students to complete the analysis.

This preference for the detailed step-by-step and branching structures is deeper than a simple desire to "just get the answer". Student use of an online tutorial is a balancing act between time, effort, and utility. When the perceived difficulty or tedium of a specific step in a tutorial becomes too great, students switch to a different resource. In CITA, this was especially apparent in numerical and check box questions bundled in the tutorials.

We recommend that researchers who design scaffolds in online environments be well versed in the expectations of their students. That way, they can tailor tutorials to fit the needs of their classes.

5.11 Acknowledgments

This work is supported in part by the College of Science at Purdue University. We would also like to thank the PER@Purdue research group for their helpful discussions and insights.

6. SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS

6.1 Summary

In this study, I described how there was a pressing need for a tutorial system to compliment CHIP - the current homework system used in Electricity and Optics. This need was caused by a variety of factors including the development of the distance learning sections of PHYS 24100, rising enrollment in summer, an aging CHIP homework system, and student feedback from previous semesters.

I created interactive tutorials that were paired with problems in the PHYS 24100 homework. The tutorials were built using three structures that took students through Polya's problem solving process: Shallow CITA, Immersive CITA, and Postscripts. The parts were not meant to be used separately. Shallow CITA provided focused feedback when students made simple errors on homework problems. If a student could not solve for the correct answer with Shallow CITA alone, he or she could use Immersive CITA to review the details of solving the problem. The Postscripts wrapped up the problem by asking followup questions or outlining similar situations where the completed analysis would apply.

6.2 Research Questions

How does the branching structure of the interactive examples in CITA influence student problem solving abilities?

Overall, the first version of CITA (summer 2015) yielded little to no gain in student performance. This is most likely due to the fact that only half the problems contained a tutorial as well as the fast pace of the class during the summer. However, versions two and three of CITA show small but consistent gains for students that use the system. Students that used the CITA system had higher course grades than those who did not. Additionally, on-campus students who used CITA had higher scores on select exam problems as well as better performance on a diagnostic multi-step problem than those who did not. Unfortunately, online students did not show the same gains.

What are students perceptions about the CITA system?

Shallow CITA was very popular with the students in PHYS 24100. They found it extremely useful for identifying simple mistakes like unit errors and sign errors that would have otherwise led to frustration with finding a solution. Students suggested that we expand these tutorials, but also mentioned that they were more detailed than some of the other homework systems that they had used in the past (like WebAssign).

Students were generally satisfied with Immersive CITA, but not all tutorial structures were equally popular. Generally speaking, students did not like the "truncated step-by-step" style of tutorial; even if it outlined the solution to the problem, they generally found it unhelpful and turned to other online sources for help. Additionally, the problems with no Immersive CITA tutorial were also disliked because they did not provide the same feedback as so many of the other questions in the homework. However, the detailed step-by-step and branching tutorials that I developed were warmly received.

Students overwhelmingly ignored the Postscripts at the end of each problem. Many of them responded that they did not consider the Postscripts due to time constraints (discussed further below). The small number of students that completed the Postscripts commented that they were helpful in learning the concepts in the problem.

Overall, the CITA system was positively received by the student body. They thought it was a good resource for completing the homework problems.

What motivates a student to use CITA rather than another external source?

Focus group interviews and click data showed that the CITA tutorials were used by many of the students in PHYS 24100. Students created a tool belt of resources that they used to complete the homework as quickly and efficiently as possible (learning the material was often secondary). Other resources that were used by students included their notes, the textbook, Yahoo! Answers, and HyperPhysics.

Students are under immense pressure to complete their coursework to the best of their abilities. Thus, physics instructors try to provide useful materials to aid students' development of understanding abstract concepts. In our case, we designed a system called CITA to help students through their homework.

Students are constantly assessing the work needed to complete the tutorial along with the content in the tutorial itself. When the perceived effort for a particular step in the tutorial exceeds the perceived benefit gained from continuing onwards, students leave the tutorial and move onto another resource. This switching might not be possible in a setting where changing resources takes a lot of effort. However, it is possible online; the Internet offers a variety of resources at the click of a button. Thus, instructors who are designing tutorial systems must consider the effort needed to navigate their tutorial system as compared with other external systems used by students in their class.

6.3 Future Directions

The development of a homework system is not something that is complete in a few iterations. In this section, I outline several future directions that other researchers can follow.

6.3.1 Longitudinal Study of CITA

Although the research group collected data from reasonably large samples of students, the fact remains that we have only calculated the effects of the final state of CITA on the spring 2016 and summer 2016 semesters. A longitudinal study of CITA, where the system is analyzed over multiple semesters, would probe the results on a deeper level. It would increase our understanding of how students navigate the tutorials as well as shed light on where in the tutorials students exit.

This information can be used to design more effective lessons in the future. The theoretical framework of constructivism dictates that students actively build a model of how the world works as they traverse a CITA tutorial. Thus, if one identifies parts of the analysis that students consistently find troublesome, then an educator can expand on the CITA tutorial at that point. This might include reviewing mathematical methods, outlining electromagnetism concepts on a deeper level, or relating ideas in Electricity and Optics back to those in introductory mechanics.

6.3.2 Create Multiple Paths with Different Tutorials

The majority of the CITA tutorials go through one type of solution per problem. The latest version of CITA allows students to choose between two paths for their tutorial, but the content of the tutorial is the same; all that changes is the pace. Although I tried to choose the most common solution to a problem that exploits relevant symmetries, there is always more than one way to arrive at the correct answer.

A researcher could explore whether providing students with different solution paths would help them learn the material more effectively. A student could tackle a problem in his or her own unique way, based on the relevant fundamental laws that he or she chooses to use. This idea fits the theoretical framework of constructivism very well; a student would actively build a complete solution to a problem, starting from fundamental laws that he or she understands. The starting point of a solution would not be dictated by a tutorial, but rather by a student's current model of the world.

6.3.3 Implement "Dead End" Style Tutorials

During the development of the branching structure within CHIP, the research group thought about problems that would use a "dead-end" style of tutorial. A student would go down a seemingly reasonable path of logic and would eventually reach a contradiction in the analysis. Then, the tutorial would outline the contradiction, take the student back to the appropriate point in the tutorial, and continue down a new path. Although I did create a proof of concept for the dead-end style of tutorial, I did not implement these widely throughout the system. I did not know how to effectively tackle a number of problems that would arise, including:

1. How does one handle a student who leaves the tutorial early and misses the highlight of the contradiction?

2. How does one handle students who are reluctant to go through a dead-end tutorial due to the larger time commitment?

3. How does one handle students who develop incorrect ideas about fundamental physics concepts because they do not carefully read through the passages that separate the true concepts from the false ones?

This structure would allow an educator to design truly interactive tutorials where students could explore physics concepts, both right and wrong. As stated in Chapter 2, understanding and working through one's mistakes is when learning takes place.

6.3.4 Conduct a Detailed Graphical User Interface Study

The research group decided early on in the development of CITA on CHIP to keep the graphical user interface (GUI) the same between semesters. Although the menus changed slightly from semester to semester, students saw the same problems in as close to the same format as possible (save for grammar corrections, updated line spacing, and refurbished diagrams). This allowed us to make comparisons between classes with consistent assignments.

However, the physical features of a multimedia system can influence its use [127]. One direction of research might examine how the GUI influences student learning gains using the CITA system as a platform. Do different color palettes improve or detract from learning? Do specific layouts of the problems help students tie together different physics concepts? A/B testing has been used to a great extent in this area in industry, but multivariate methods might also be useful in this field. This type of study would probe specific implementations of Mayer's design principles in order to determine what kind of design is most useful to teach physics.

6.4 Implications of the Research

Even though the Internet has pervaded our lives in a myriad of ways, it is still a very new technology. Researchers are still learning about the impact that it has on the way students think, learn, and live. The very nature of the internet, where information is available at the touch of a button, is unlike anything that we have ever seen.

This study has shown that online systems may be used to teach physics if the instructor designs scaffolded tutorials that walk through the full problem-solving process. Some of the structures that might be used in a live setting do not seem to work as well in the online environment (like the truncated step-by-step tutorials). Other styles of tutorials work very well because they take advantage of the structure of how web pages work and constantly build off of feedback of the user (like the branching style). These results can serve as a guide for educators who design online tutorials in the future.

LIST OF REFERENCES

LIST OF REFERENCES

[1] John P. Bean and Barbara S. Metzner. A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55(4):485–540, 1985.

[2] Laura J. Horn and C. Dennis Carroll. Nontraditional undergraduates: Trends in enrollment from 1986 to 1992 and persistance and attainment among 1989-90 beginning postsecondary studies. *NCES 97-578*, November 1996.

[3] Susan Choy. Nontraditional undergraduates. *NCES 2002-012*, 2002.

[4] Ian D. Beatty. Transforming student learning with classroom communication systems. *ECAR Research Bulletin*, 2005.

[5] Carmen Fies and Jill Marshall. Classroom response systems: A review of the literature. *Journal of Science Education and Technology*, 15(1):101–109, 2006.

[6] Jim Ong and Sowmya Ramachandran. Intelligent tutoring systems: the what and the how. *Learning Circuits*, 1(2), 2000.

[7] John Self. Theoretical foundations for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, 1(4):3–14, 1990.

[8] John R Anderson, C Franklin Boyle, and Brian J Reiser. Intelligent tutoring systems. *Science(Washington)*, 228(4698):456–462, 1985.

[9] Purdue University. Plans of study, 2016. [Online; updated 2015].

[10] P.A. Tipler and G. Mosca. *Physics for Scientists and Engineers*. Number pts. 1-33 in Physics for Scientists and Engineers. W. H. Freeman, 2007.

[11] Cisco webex, 2016. [Online; updated August-2016].

[12] Ruth Chabay and Bruce Sherwood. *Matter and Interactions*, volume 1. John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2010.

[13] V.K. Saxena, H. Nakanishi, D. Elmore, D. Shibata, and D. Bortoletto. Chip: Computerized homework in physics. *Using Technology to Enhance Learning: How Does It Change What Faculty Do?*, 1998.

[14] Hisao Nakanishi and V.K. Saxena. Computerized homework in physics, 2016. [Online; updated August-2016].

[15] Paul A. Tipler and Gene Mosca. *Physics for Scientists and Engineers: Standard Version*. W.H. Freeman, 5 edition, July 2003.

[16] Alan Giambattista, Betty Richardson, and Robert Richardson. *Physics*. McGraw-Hill Education, 2 edition, January 2009.

[17] David Halliday, Robert Resnick, and Jearl Walker. *Fundamentals of Physics*. Wiley, 10 edition, August 2013.

[18] John D. Cutnell and Kenneth W. Johnson. *Introduction to Physics*. John Wiley and Sons Ltd., 8 edition, April 2009.

[19] Advisory Committee to the National Science Foundation. Shaping the future: Perspectives on undergraduate education in science, mathematics, engineering, and technology. 2, 1996.

[20] American Educational Research Association, 2016. [Online; updated August-2016].

[21] Ernst von Glasersfeld. Learning as constructive activity. *Proceedings of the 5th Annual Meeting of the North American Group of Psychology in Mathematics Education*, 1, 1983.

[22] Jeanne Ellis Ormrod. *Human Learning*. Pearson, 6 edition, September 2011.

[23] Ernst von Glasersfeld. *Constructivism in Education*, volume 1. Pergamon Press, 1989.

[24] Ernst Von Glasersfeld. Cognition, construction of knowledge, and teaching. *Synthese*, 80(1):121–140, 1989.

[25] Ernst von Glasersfeld. *Aspects of Radical Constructivism (Translated from Aspectos del constructivismo radical)*. Gedisa Editorial, 1996.

[26] Jean Piaget and Barbel Inhelder. *The Psychology of the Child*. Basic Books, 2 edition, October 1969.

[27] Jean Piaget. *The Child's Conception of the World: A 20th-Century Classic of Child Psychology*. Littlefield Adams Quality Paperbacks, January 1975.

[28] Jean Piaget. *The Psychology of Intelligence*. Routledge, 2 edition, June 2001.

[29] L.S. Vygotsky et al. *Thought and Language - Revised Edition*. The MIT Press, August 1986.

[30] Lev S. Vygotsky and Alex Kozulin. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, March 1978.

[31] Jerome Bruner. The act of discovery. *Harvard Educational Review*, 31(1), 1961.

[32] Jerome S. Bruner. *Toward a theory of instruction*. Belknap Press of Harvard University Cambridge, Mass, 1966.

[33] Jerome S. Bruner and Jeremy M. Anglin. *Beyond the Information Given: Studies in the Psychology of Knowing*. W. W. Norton and Company, May 1973.

[34] Jerome Bruner. *The Process of Education*. Harvard University Press, September 1977.

[35] Daniel R. Miller. *American Sociological Review*, 31(1):128–130, 1966.

[36] A. Bandura. *Principles of behavior modification*. Holt, Rinehart, and Winston, 1 edition, July 1969.

[37] Albert Bandura. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191, 1977.

[38] Albert Bandura. *Social Learning Theory*. Prentice-Hall, 1 edition, November 1976.

[39] George Polya. *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press, 2 edition, 1985.

[40] Alan H Schoenfeld. *Mathematical problem solving*. Elsevier, 2014.

[41] D.P. Maloney. *Handbook of research on science teaching and learning*. Macmillan, New York, NY, 1994.

[42] Laura N. Walsh, Robert Howard, and Brian Bowe. Phenomenographic study of students problem solving approaches in physics. *Phys. Rev. ST Phys. Educ. Res.*, 3(2):020108, December 2007.

[43] Edward F Redish, Rachel E Scherr, and Jonathan Tuminaro. Reverse-engineering the solution of a "simple" physics problem: Why learning physics is harder than it looks. *The Physics Teacher*, 44(5):293–300, 2006.

[44] John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[45] Jill Larkin, John McDermott, Dorothea P. Simon, and Herbert A. Simon. Expert and novice performance in solving physics problems. *Science*, 208(4450):1335–1342, 1980.

[46] Michelene TH Chi, Robert Glaser, and Ernest Rees. Expertise in problem solving. 1981.

[47] Michelene TH Chi, Paul J Feltovich, and Robert Glaser. Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2):121–152, 1981.

[48] Martha Larkin. Using scaffolded instruction to optimize learning. *ERIC Digest*, December 2002.

[49] Lisbeth Dixon-Krauss. *Vygotsky in the Classroom: Mediated Literacy Instruction and Assessment*. Pearson, 1 edition, November 1995.

[50] Robert Slavin. *Educational Psychology: Theory and Practice*. 10 edition, January 2011.

[51] W. Jung. Uses of cognitive science to science education. *Science & Education*, 2(1):31–56, 1993.

[52] Antonio Neto and Maria Odete Valente. Problem solving in physics: Towards a metacognitively developed approach. 1997.

[53] Phillip Dukes, David E Pritchard, and Elsa-Sofia Morote. Inductive influence of related quantitative and conceptual problems. 2002.

[54] William J. Leonard, Robert J. Dufresne, and Jose P. Mestre. Using qualitative problem solving strategies to highlight the role of conceptual knowledge in solving problems. *American Journal of Physics*, 64(12):1495–1503, 1996.

[55] University of Minnesota PER Group. C3po: Customizable computer coaches for physics, 2016. [Online; updated August-2016].

[56] Patricia Heller and Mark Hollabaugh. Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7):637–644, 1992.

[57] ERIC MAZUR. The problem with problems. *Optics and photonics news*, 7(6):59–60, 1996.

[58] Alan Van Heuvelen and David P. Maloney. Playing physics jeopardy. *American Journal of Physics*, 67(3):252–256, 1999.

[59] Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, 2 edition, January 2009.

[60] John Sweller. *Instructional Design in Technical Areas*. Australian Council for Education, April 1999.

[61] Richard Mayer. Research-based principles for multimedia learning, 2014. Harvard Initiative for Learning and Teaching (HILT).

[62] Christopher A Sanchez and Jennifer Wiley. An examination of the seductive details effect in terms of working memory capacity. *Memory & cognition*, 34(2):344–355, 2006.

[63] Laura E Thomas and Alejandro Lleras. Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review*, 14(4):663–668, 2007.

[64] Adrian Madsen, Amy Rouinfar, Adam M. Larson, Lester C. Loschky, and N. Sanjay Rebello. Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Phys. Rev. ST Phys. Educ. Res.*, 9:020104, Jul 2013.

[65] Amy Rouinfar, Elise Agra, Jeffrey Murray, Adam Larson, Lester C. Loschky, and N. Sanjay Rebello. Influence of visual cueing on students' eye movements while solving physics problems. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, pages 191–194, New York, NY, USA, 2014. ACM.

[66] Ruth C. Clark and Richard E. Mayer. *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. Wiley, 4 edition, March 2016.

[67] Amy Marie Oberfoell. Understanding the role of the modality principle in multimedia learning environments. 2015.

[68] Microsoft Corporation. Dictate text using speech recognition. https://support.microsoft.com/en-us/help/14198/windows-7-dictate-text-using-speech-recognition.

[69] Apple Corporation. Official apple support. https://support.apple.com.

[70] Pocket sphinx package. https://launchpad.net/ubuntu/+source/pocketsphinx.

[71] Robert Dufresne, Jose Mestre, David M Hart, and Kenneth A Rath. The effect of web-based homework on test performance in large enrollment introductory physics courses. *Journal of Computers in Mathematics and Science Teaching*, 21(3):229–252, 2002.

[72] Michelle Richards-Babb, Janice Drelick, Zachary Henry, and Jennifer Robertson-Honecker. Online homework, help or hindrance? what students think and how they perform. *Journal of College Science Teaching*, 40(4):81–93, 2011.

[73] Jane Dillard-Eggers, Tommy Wooten, Brad Childs, and John Coker. Evidence on the effectiveness of on-line homework. *College Teaching Methods & Styles Journal (CTMS)*, 4(5):9–16, 2011.

[74] Webassign, 2016. [Online; updated August-2016].

[75] Mcgraw hill digital platforms, 2016. [Online; updated August-2016].

[76] Sapling learning, 2016. [Online; updated August-2016].

[77] Sapling learning, 2016. [Online; updated August-2016].

[78] Mastering physics, 2016. [Online; updated August-2016].

[79] The expert ta, 2016. [Online; updated August-2016].

[80] Lon-capa, 2016. [Online; updated August-2016].

[81] Flipit physics, 2016. [Online; updated August-2016].

[82] G.M. Novak. *Just-in-time Teaching: Blending Active Learning with Web Technology*. Ellis Horwood Series in Environmental Management, Science an. Prentice Hall, 1999.

[83] Eugenia Etkina, Anna Karelina, Maria Ruibal-Villasenor, David Rosengrant, Rebecca Jordan, and Cindy E Hmelo-Silver. Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories. *The Journal of the Learning Sciences*, 19(1):54–98, 2010.

[84] Minds on physics modules, 2016. [Online; updated August-2016].

[85] Blackboard learn, 2016. [Online; updated August 2016].

[86] Moodle, 2016. [Online; updated August-2016].

[87] Moodle rooms, 2016. [Online; updated August-2016].

[88] Andreas M Kaplan and Michael Haenlein. Higher education and the digital revolution: About moocs, spocs, social media, and the cookie monster. *Business Horizons*, 2016.

[89] Khan academy, 2016. [Online; updated August-2016].

[90] Udacity, 2016. [Online; updated August-2016].

[91] edx, 2016. [Online; updated August-2016].

[92] Coursera, 2016. [Online; updated August-2016].

[93] Academic earth, 2016. [Online; updated August-2016].

[94] Abbas Tashakkori and Charles Teddlie. *Handbook of mixed methods in the social and behavioral sciences*. Thousand Oaks, CA: Sage, 2003.

[95] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 2 edition, 1988.

[96] H. J. Keselman Lisa M. Lix, Joanne C. Keselman. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "f" test. *Review of Educational Research*, 66(4):579–619, 1996.

[97] Alan Agresti. *An introduction to categorical data analysis*. Wiley, John, & Sons Inc., New York, 1 edition, February 1996.

[98] Kevin R Murphy, Brett Myors, and Allen Wolach. *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Routledge, 2014.

[99] Barbara G Tabachnick, Linda S Fidell, and Steven J Osterlind. Using multivariate statistics. 2001.

[100] Ruth Chabay and Bruce Sherwood. Brief electricity and magnetism assessment, 1997. http://www.compadre.org/per/items/detail.cfm?ID=3775.

[101] Lin Ding, Ruth Chabay, Bruce Sherwood, and Robert Beichner. Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Phys. Rev. ST Phys. Educ. Res.*, 2:010105, Mar 2006.

[102] Lin Ding. Seeking missing pieces in science concept assessments: Reevaluating the brief electricity and magnetism assessment through rasch analysis. *Phys. Rev. ST Phys. Educ. Res.*, 10:010105, Feb 2014.

[103] Bruce Sherwood. Tracking steps in multistep problems. AAPT Summer Meeting 2005, August 2005.

[104] Ruth Chabay and Bruce Sherwood. Thinking iteratively. AAPT Summer Meeting 2014, 2014.

[105] J. Saldana. *The Coding Manual for Qualitative Researchers*. SAGE Publications, 2012.

[106] Monique Hennink. *Focus Group Discussions: Understanding Qualitative Research*. Oxford University Press, 1 edition, January 2014.

[107] Michael Quinn Patton. *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*. SAGE Publications Inc., 2455 Teller Road, Thousand Oaks, CA 91320, 4 edition, November 2015.

[108] Norman Denzin and Yvonna S. Lincoln. *The Landscape of Qualitative Research*. SAGE Publications, Inc, 2455 Teller Road, Thousand Oaks, CA 91320, 4 edition, 2012.

[109] Piazza, 2016. [Online; updated August-2016].

[110] J.W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2003.

[111] Rev.com. Accessed July 2016.

[112] Giambattista Vico. *New Science*. Penguin, 3 edition, Original work published 1725.

[113] John Dewey. *Experience and education*. Kappa Delta Pi, 1998.

[114] Georg Wilhelm Friedrich Hegel. *The phenomenology of mind*. Courier Corporation, 2012.

[115] Catherine Twomey Fosnot. *Constructivism: Theory, perspectives, and practice*. Teachers College Press, 2 edition, August 2013.

[116] Adriann DeGroot. *Thought and Choice in Chess (Psychological Studies)*. 2 edition, 1978.

[117] Graham Davey. *Encyclopaedic Dictionary of Psychology*, volume 1. Routledge, July 2006.

[118] Gerd Kortemeyer. Gender differences in the use of an online homework system in an introductory physics course. *Physical Review Special Topics-Physics Education Research*, 5(1):010107, 2009.

[119] Laura McCullough. Gender differences in student responses to physics conceptual questions based on question context. *ASQ Advancing the STEM Agenda in Education*, 2011.

[120] S Güzin Mazman and Yasemin Koçak Usluel. Gender differences in using social networks. *TOJET: The Turkish Online Journal of Educational Technology*, 10(2), 2011.

[121] Noah S. Podolefsky and Noah D. Finkelstein. Analogical scaffolding and the learning of abstract ideas in physics: An example from electromagnetic waves. *Phys. Rev. ST Phys. Educ. Res.*, 3:010109, Jun 2007.

[122] Christine Lindstrm and Manjula Sharma. Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter? *Int. J. Innov. Sci. Math. Educ.*, 19(2):1–19, 2011.

[123] Elisheva Cohen, Andrew Mason, Chandralekha Singh, and Edit Yerushalmi. Identifying differences in diagnostic skills between physics students: Students' self-diagnostic performance given alternative scaffolding. *arXiv preprint arXiv:1603.03105*, 2016.

[124] Arthur C Graesser, Danielle S McNamara, and Kurt VanLehn. Scaffolding deep comprehension strategies through point&query, autotutor, and istart. *Educational psychologist*, 40(4):225–234, 2005.

[125] A Stinner. Providing a contextual base and a theoretical structure to guide the teaching of high school physics. *Physics Education*, 29(6):375, 1994.

[126] Ch Chen. An adaptive scaffolding e-learning system for middle school students physics learning. *Australasian Journal of Educational Technology*, 30(3):342–355, 2014.

[127] Hyungjoo Park and Hae-Deok Song. Make e-learning effortless! impact of a redesigned user interface on usability through the application of an affordance design approach. *Educational Technology & Society*, 18(3):185–196, 2015.

[128] Robert J. Beichner. Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8):750–762, 1994.

[129] David Hestenes, Malcolm Wells, and Gregg Swackhamer. Force concept inventory. *Phys. Teach.*, 30(3):141–158, March 1992.

[130] Ronald K. Thornton and David R. Sokoloff. Assessing student learning of newtons laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, 66(4):338–352, 1998.

[131] Paula Engelhardt and Robert Beichner. Students' understanding of direct current resistive electrical circuits. *Am. J. Phys.*, 72(1):98–115, January 2004.

[132] David P. Maloney, Thomas L. OKuma, Curtis J. Hieggelke, and Alan Van Heuvelen. Surveying students conceptual knowledge of electricity and magnetism. *American Journal of Physics*, 69(S1):S12–S23, 2001.

[133] Chandralekha Singh. Student understanding of symmetry and gausss law of electricity. *American journal of physics*, 74(10):923–936, 2006.

[134] Mark F Masters and Timothy T Grove. Active learning in intermediate optics through class tutorials and concept building laboratories. In *Eleventh International Topical Meeting on Education and Training in Optics and Photonics*, pages 96660K–96660K. International Society for Optics and Photonics, 2009.

[135] Erin M Bardar, Edward E Prather, Kenneth Brecher, and Timothy F Slater. Development and validation of the light and spectroscopy concept inventory. *Astronomy Education Review*, 5(2):103–113, 2006.

[136] RJ Roedel, S El-Ghazaly, Teri Reed Rhoads, and E El-Sharawy. The wave concepts inventory-an assessment tool for courses in electromagnetic engineering. In *Frontiers in Education Conference, 1998. FIE'98. 28th Annual*, volume 2, pages 647–653. IEEE, 1998.

[137] Rebecca Lindell and Lin Ding. Establishing reliability and validity: An ongoing process. In *Physics Education Research Conference 2012*, volume 1513 of *PER Conference Invited Paper*, pages 27–29, Philadelphia, PA, August 1-2 2012.

[138] Gordon J. Aubrecht and Judith D. Aubrecht. Constructing objective tests. *American Journal of Physics*, 51(7):613–620, 1983.

[139] Scott Bonham. Reliability, compliance, and security in web-based course assessments. *Phys. Rev. ST Phys. Educ. Res.*, 4:010106, Apr 2008.

[140] Adrian Madsen and Sam McKagan. Aapt physport, August 2016. [Online; posted 12-August-2016].

[141] L. Crocker and J. Algina. *Introduction to Classical and Modern Test Theory*. Cengage Learning, 2006.

[142] Rodney L Doran. *Basic Measurement and Evaluation of Science Instruction*. ERIC, 1980.

[143] Glenn Fulcher and Fred Davidson. *Language testing and assessment*. Routledge London, England & New York, NY, January 2007.

[144] G. Henning. *A guide to language testing: development - evaluation - research*. Newbury House, Cambridge, MA, 1987.

[145] George A. Ferguson. On the theory of test discrimination. *Psychometrika*, 14(1):61–68, 1949.

[146] Paul Kline. *A Handbook of Test Construction (Psychology Revivals): Introduction to Psychometric Design*. Routledge, 2015.

[147] G. F. Kuder and M. W. Richardson. The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160, 1937.

[148] James L Bruning and Buddy L Kintz. *Computational Handbook of Statistics*. Scott, Foresman & Co, 3 edition, April 1987.

APPENDICES

A. CHEATING ON THE BEMA CONCEPT INVENTORY

A.1 Preface

During our analysis of the CITA on CHIP system, we noticed some anomalies between the BEMA scores in the online and on-campus sections. The online scores seemed far too high to be reasonable for students at an introductory level. Thus, we performed a comparison of the BEMA scores between the online and on-campus sections, taking inspiration from Lin Ding's earlier analysis of the BEMA [101, 102]. This analysis was published in the paper shown below. By the time this dissertation is published, our paper on the BEMA will most likely be in pre-print.

A.2 Introduction

Recently, a number of concept inventories have been developed to assess the learning gains of students in their introductory physics courses. These tests cover a wide range of topics including classical mechanics [128–130], electromagnetism [131–133], and optics [134–136] (among many others). The Brief Electricity and Magnetism Assessment [100] (BEMA) was developed by Chabay, Sherwood, and Reif to measure student's qualitative understanding of electricity and magnetism concepts at the introductory level. The full exam can be found on the Compadre website at http://www.compadre.org.

Although concept inventories are a useful method of probing student knowledge, they are not simply tests that can be quickly put together and administered semester after semester. Lindell and Ding describe how it takes years to determine the validity and reliability of the results of a given concept inventory [137]. Reliability is a measure of consistency. If a test is taken several times under similar conditions and the results are similar, then the test is considered to be reliable. Validity is a measure of accuracy. A test is considered valid if it accurately measures what it claims to measure. It is important to note that reliability and validity are two separate concepts that are independent of each other [138]. Factors such as age, course structure, geography, language, delivery of the tests, and wording of questions can influence the validity and reliability of an assessment.

Online environments have the additional challenge of being unsupervised. Unlike a classroom setting where a teacher can monitor students to check for cheating, online students can take exams in a private location. This means that there is an increased chance of sharing answers, referencing external sources, and using unapproved equipment (e.g. graphing calculators). To make matters worse, if the assessment does not have a point value, many students may skip it altogether [139]. Thus, the AAPT has issued a set of guidelines that are recommended for any instructor who wants to administer a concept inventory online [140].

Much work has been done to validate the BEMA in an on-campus setting by Lin Ding and his team [101, 102]. However, an analysis of the reliability, validity, and discriminatory power of the BEMA has not been done for an online setting. Our research questions are:

1. Is the BEMA a reliable and valid test in an online setting with sufficient discriminatory power between students?
2. If the BEMA is not reliable and/or valid in online settings, how do different measures of reliability, validity, and discriminatory power differ between the online and on-campus environments?

There is a pressing need to answer these questions due to the wealth of information online along with the growing trend of administering assessments online. With this information, educators can make predictions about who might be a potential risk for failure or dishonest behavior on the BEMA.

A.3 Background

A.3.1 Overview of the Classes

Our study took place in the calculus-based, introductory electromagnetism courses at Purdue University (coded as PHYS 24100 and PHYS 24100D). These courses are usually taken by sophomore engineering majors as a prerequisite for their later engineering classes (the exception being electrical engineering majors who take PHYS 27200) [9]. Both PHYS 24100 and PHYS 24100D are titled "Electricity and Optics". The "D" simply signifies an online section of the course in the fall and spring semesters. During the summer semester, the course is only administered online, so the separate PHYS 24100D code is not necessary.

The content in the on-campus and online sections of Electricity and Optics is exactly the same. All of the students use *Physics for Scientists and Engineers* by Tipler and Mosca as their textbook [10]. The one-semester course covers chapters 21 through 33, with topics including: electric charge, electric fields, electric potential, circuits, magnetic fields, magnetic induction, Maxwell's equations, geometric optics, and interference effects. The online and on-campus sections only differ in the fact that the online sections watch pre-recorded video lectures and attend online recitations through Cisco WebEx [11] while the on-campus sections attend live lectures and live recitations.

A.3.2 Class Demographics

Engineering students usually take this course in the fall semester (the "on-semester"). Alternatively, some engineering students choose to take the course in the spring or summer (the "off-semesters") due to the fact that they are either ahead or behind in their schedules. Table A.1 shows the total enrollment in the campus and online sections of Electricity and Optics over the last two years. During the fall and spring semesters, just under 10% of the total class is enrolled in PHYS 24100D.

Table A.1
Enrollment in the campus and online sections of Electricity and Optics from spring 2015 through spring 2016. All of the enrollment numbers ignore dropped students.

Semester	**Online**	**Campus**	**Percent Online**
Spring 2015	43	474	8.3%
Summer 2015	187	0	100.0%
Fall 2015	69	728	8.7%
Spring 2016	36	339	9.6%

Before analyzing any of the BEMA scores, we wanted to compare the physics and mathematics abilities of the students in each section. All of the students in PHYS 24100 were asked to complete a demographic survey at the beginning of the semester that asked for information on their previous calculus and physics classes. Each of the questions on the survey included an option of "I Prefer Not to Disclose", just in case a student wanted to keep their information private. However, the vast majority of the students answered all of the demographics questions (just under 99%).

Two of the questions on the survey asked students about their previous grades in their introductory mechanics class and their first semester calculus class (if they were taken at Purdue University). We found that the student bodies had very different mean scores between the on-semester (fall 2015) and the off-semesters (spring 2015, summer 2015, and spring 2016). However, the online and on-campus sections within a single semester had very similar scores in each case. Tables A.2 and A.3 show the mean mechanics and calculus grades in each of the semesters. Though some of the mean grades on these tables appear quite distinct at first sight, statistical tests indicate otherwise.

Chi-squared tests of the data showed no statistically significant differences between the online and on-campus grade distributions within a semester except for the fall

Table A.2
Reported mean mechanics grades for students in PHYS 24100. Students were asked to report their grade in their previous mechanics course (with the option of not disclosing this information should they want to keep it private). The majority of students took PHYS 17200 - the introductory mechanics course at Purdue University. The options were "A" (4), "B" (3), "C" (2), and "D" (1).

Semester	**Mean (Campus)**	**Mean (Online)**
Spring 2015	2.726	2.333
Summer 2015	NA	2.531
Fall 2015	3.022	2.795
Spring 2016	2.871	2.556

Table A.3
Reported mean calculus grades for students in PHYS 24100. Students were asked to report their grade in their most recent calculus course (with the option of not disclosing this information should they want to keep it private). Most students had completed the second semester of introductory calculus and were taking Calculus III concurrently with PHYS 24100 or 24100D. The options were "A" (4), "B" (3), "C" (2), and "D" (1).

Semester	**Mean (Campus)**	**Mean (Online)**
Spring 2015	3.091	2.893
Summer 2015	NA	3.055
Fall 2015	3.255	3.065
Spring 2016	3.121	3.136

2015 and spring 2016 calculus comparisons (see Table A.4). However, the effect sizes (measured by Cramer's V) for the fall of 2015 and spring of 2016 are 0.139 and 0.193 respectively. We will use the traditional cutoff of $p = 0.05$ for statistical significance and Cohen's guidelines for interpreting effect sizes in this paper [95], so we conclude that they are relatively small effect sizes. Since the grade distributions were similar between the online and on-campus sections, we would expect the initial BEMA grade distributions to be similar as well. However, as we will see, this was definitely not the case.

Table A.4
The p-values of the chi-squared tests comparing scores in the online and on-campus sections. We only compare grades of A, B, and C since the number statistics were very low for D grades. Wherever there was a statistically significant difference in the distributions, we saw a small effect size.

Semester	**Mechanics**	**Calculus**
Spring 2015	0.124	0.096
Fall 2015	0.085	0.032
Spring 2016	0.342	0.030

A.3.3 Administration of the BEMA

The BEMA was administered twice during each semester as a pre-test and post-test. In the spring and fall semesters, the pre-test was given during the first week of class while the post-test was given around the 12th week (right after the students had finished the unit on magnetic induction and were moving onto AC circuits). During the summer, the BEMA pre-test and post-test were administered during the first and sixth weeks of the semester.

The on-campus sections took the BEMA exam during their scheduled recitation periods while the online sections took the exam through our in-house online homework system called CHIP (Computerized Homework in Physics). In both cases, the students could see the entire exam at one time and answer the questions in any order that they wished. As per the suggested guidelines mentioned in the introduction, the online students were told that they were taking a “diagnostic exam”, and they would receive credit for one recitation quiz if they tried their best. Both the online and on-campus sections were given a time limit of 30 minutes to complete the BEMA. None of the students were able to see their final scores on the BEMA; instead, they only saw a recitation quiz credit of 10 points if they attempted the test.

The inspiration for this study came from viewing the histograms of BEMA pre-test and post-test scores (see Figure A.1). There is a very obvious discrepancy between students in the on-campus sections (where the distributions of scores look like slightly skewed normal distributions) and the students in the online sections (where the distributions of scores have a very prominent spike around a perfect score). Interestingly, the abnormal peak in the online sections does not occur at a score of 31/31; rather, it occurs at a score of 30/31. The reason for this is explained in Section A.6.2.

A.4 Methods of Analysis

A.4.1 Measures of Reliability and Discrimination

To identify whether online students are experiencing the BEMA differently than their on-campus counterparts we compare a number psychometric properties of the test, dis-aggregated by group. Pulling from Lin Ding's earlier assessment of the BEMA [101], we primarily focus on classical test theory (CTT) measures. These measures are based off the performance of all students on the assessment, so notable differences in the measures between the campus and online sections form the basis for claims that the two sections are not experiencing the test in an equivalent manner. The data for this analysis comes from the four semesters of PHYS 24100 and PHYS 24100D described above.

We also compare BEMA scores with exam performance to identify scores that are outliers. While the online students do take the BEMA online, they are required to take the exams on campus during the fall and spring semesters. Using the exams as an baseline measure of students ability levels we compare their overall exam score

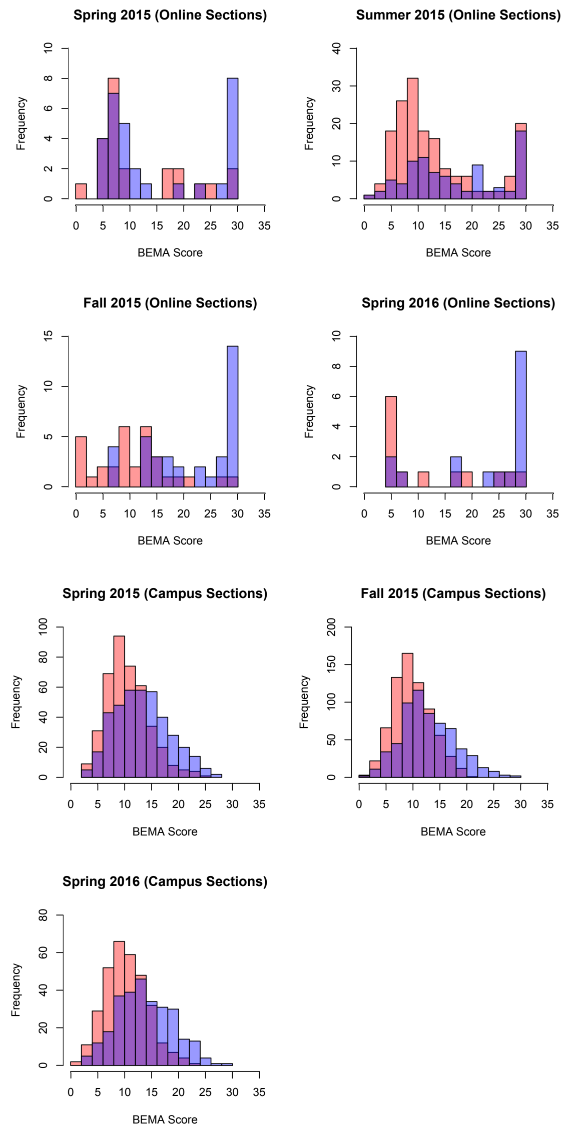

Figure A.1. Distribution of BEMA scores for the online and on-campus sections of Electricity and Optics. The red bars represent the pre-test scores while the blue bars represent the post-test scores. The on-campus sections follow a normal distribution with a slight skew (as one might expect). However, the online sections have an abnormal peak at a score of 30/31. Note that there were no on-campus sections in the summer of 2015.

with their overall BEMA scores. We then compare similar questions on the BEMA and the exams to demonstrate further inconsistencies between the groups.

A.4.2 Roles of the Researchers

In this study the principle investigators are Mr. Cyrus Vandrevala, Mr. Gary Johns, Dr. Lynn Bryan, Dr. Andrew Hirsch, Dr. Hisao Nakanishi, and Dr. Laura Pyrak-Nolte. Laura Pyrak-Nolte is the head instructor of PHYS 24100 and 24100D while Hisao Nakanishi manages the online CHIP homework system (including the setup of the BEMA in each semester). With their guidance, Cyrus Vandrevala and Gary Johns prepared the measures of the BEMA data.

Drs. Lynn Bryan and Andrew Hirsch provided the educational theory that supports this study. Neither of them are directly involved with the day to day affairs of PHYS 24100 or PHYS 24100D. It should be noted that Andrew Hirsch is one of the main instructors of PHYS 17200 - the introductory physics class that is a prerequisite for PHYS 24100 and PHYS 24100D.

A.5 Results

A.5.1 Classical Test Theory

Classical test theory (CTT) includes a number of useful measures that can be calculated for all of the assessment items, including the item difficulty index, the item discrimination index, and the point-biserial correlation coefficient. Whole test measures, such as Ferguson's Delta and the Kuder-Richardson Reliability Index (KR-21) provide additional information about a properly applied assessment [141]. We are looking at how the BEMA fails to work in an online environment.

To compare the the online and on-campus sections we look both at the differences in an item measure between the two groups at the time of pre- and post-testing as well as the changes in that measure for of each group between pre- and post-testing.

Item Difficulty Index

The item difficulty index (P) in CTT is defined as the number of correct responses to a single test question (N_c) divided by the total number of students who attempted the question (N_{tot}).

$$P = \frac{N_c}{N_{tot}} \tag{A.1}$$

This proportion varies from 0 to 1 with ideal values for items dependent on the purpose of the assessment. A difficulty index of zero means that nobody got the question correct while a difficulty index of one indicates that every test taker got the question correct.

Figure A.2 shows the trends in item difficulty for the online and campus sections between pre- and post-testing. Given that the course content covers most of the topics on the BEMA, we expect a positive correlation between the pre-test and post-test scores. The intercepts of linear regression and the clustering of values are the important distinctions between the two groups.

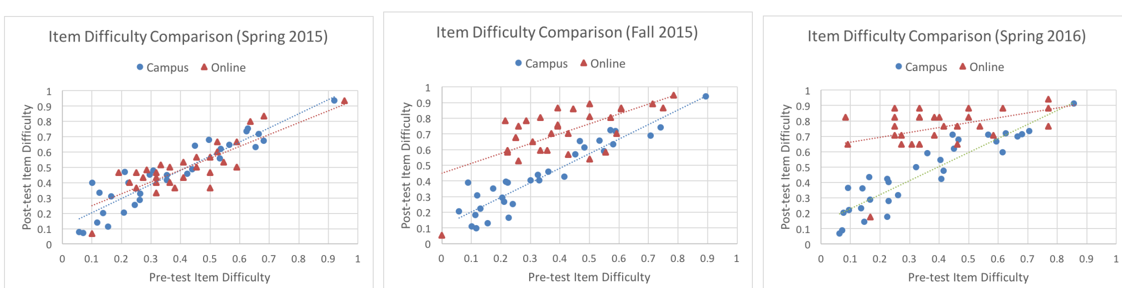

Figure A.2. Item difficulties for online and on-campus students of PHYS 24100. Each point refers to a question on the BEMA.

With the exception of a single outlier (which we will discuss later) students in the online classes see generally higher item difficulties than their campus counterparts. This means that the on-campus students are experiencing select problems as more difficult than their online colleagues.

Figure A.2 shows a particularly pronounced manifestation of this effect. Not only are the pre-test item difficulties higher for the online students, but those students also have extremely high item difficulties in the post-test; post-test scores have a lower bound of 0.6 in the online sections of spring 2016 (again, ignoring the one outlier).

To further compare the two groups, Figure A.3 shows the trend in item difficulty *between* groups for both the pre- and post-tests in the fall 2015 and spring 2016 semesters. If we were sampling truly equivalent populations of students we would expect similar item difficulties on a single test. However, the item difficulties are generally higher for the online students especially on the post test. The spring 2015 semester does not clearly separate the two groups.

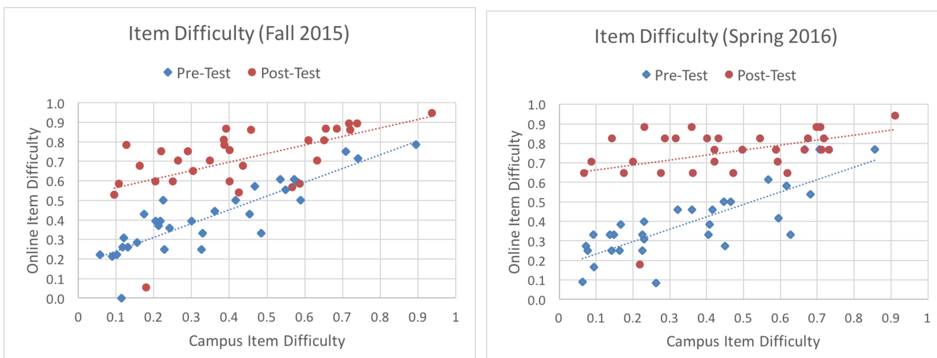

Figure A.3. Comparison of pre-test and post-test item difficulties for the online and on-campus sections. The graph on the left shows the data in the fall 2015 semester while the graph on the right shows the data from the spring 2016 semester.

When looking at the average item difficulty of all questions on the BEMA, once again we see that the online sections outperform the campus sections. Table A.5 compares the average item difficulties between the online and campus sections of PHYS 24100.

Based on the item difficulty measure, we observe that the online students are experiencing the assessment differently than their on-campus counterparts. A majority of the questions appear to be easier for the online students than their on-campus

Table A.5
Average item difficulties on the BEMA. The data below is for the spring 2015 (SP15), summer 2015 (SU15), fall 2015 (FA15), and spring 2016 (SP16) semesters.

Pre-Test Item Difficulties				
	SP15	**SU15**	**FA15**	**SP16**
Campus	0.364	NA	0.345	0.352
Online	0.424	0.454	0.414	0.394
Post-Test Item Difficulties				
	SP15	**SU15**	**FA15**	**SP16**
Campus	0.448	NA	0.431	0.394
Online	0.502	0.574	0.711	0.755

counterparts. Additionally, we see that the online group's behavior is inconsistent between semesters.

Item Discrimination Index

The item discrimination index (D) measures the discriminatory power of each item in an assessment. A test item with a high discrimination index implies that students with greater knowledge of the material will probably answer the question correctly, while students with less knowledge of the material will probably answer the question incorrectly. Conversely, a poorly constructed test question with a low item discrimination index would cause students who understand the material to answer incorrectly. Ideally, a test contains many questions with large item discrimination indices so that students with more knowledge of the material are distinguished from those with less knowledge of the material.

The calculation of the item discrimination index consists of a few steps. First, we need to split the sample of students into two groups based on their overall BEMA score. One group consists of students in the top X percent of the test scores while the other group consists of students in the bottom X percent of the test scores. The value of X is determined by the researcher, but values of 50%, 33%, and 25% are common. Then, number of students who answered a given test item correctly is counted for each of the groups (N_H and N_L standing for "high" and "low"). Finally, the difference of the counts is divided by the number of test takers in each group (N). In our case, we chose a value of X = 1/3. If we were to pick X = 1/4 like in Lin Ding's analysis, the number of data points in each group would have been very small. The item discrimination index becomes:

$$D = \frac{N_H - N_L}{N/3} \tag{A.2}$$

The item discrimination index can range from 1 to +1. Generally speaking, any questions with an item discrimination index below zero should be discarded and a minimum item difficulty index of 0.3 is preferred [142]. It should be noted that test items with an index between 0 and 0.3 are not necessarily bad, but they are places where revised test questions might be substituted.

Figure A.4 shows the item discrimination indices of each question on the BEMA for the online and on-campus sections of Electricity and Optics. First and foremost, it is important to note that the apparent grid-like patterns of the online item discrimination indices are an artifact of lower than ideal number statistics in the online sections. They should not be interpreted as an underlying pattern for the item discrimination index. However, in all four semesters, we see that the on-campus item discrimination indices are generally smaller than those in the online sections (with the exception of one outlying data point in the lower left corner of each graph).

Figure A.5 clearly shows the difference between the BEMA scores in the online and on-campus sections. Online sections of the class tend to display larger item discrimination indices than on-campus sections in both the pre-test and the post-test. This means that students in each of the sections are not experiencing the same exam.

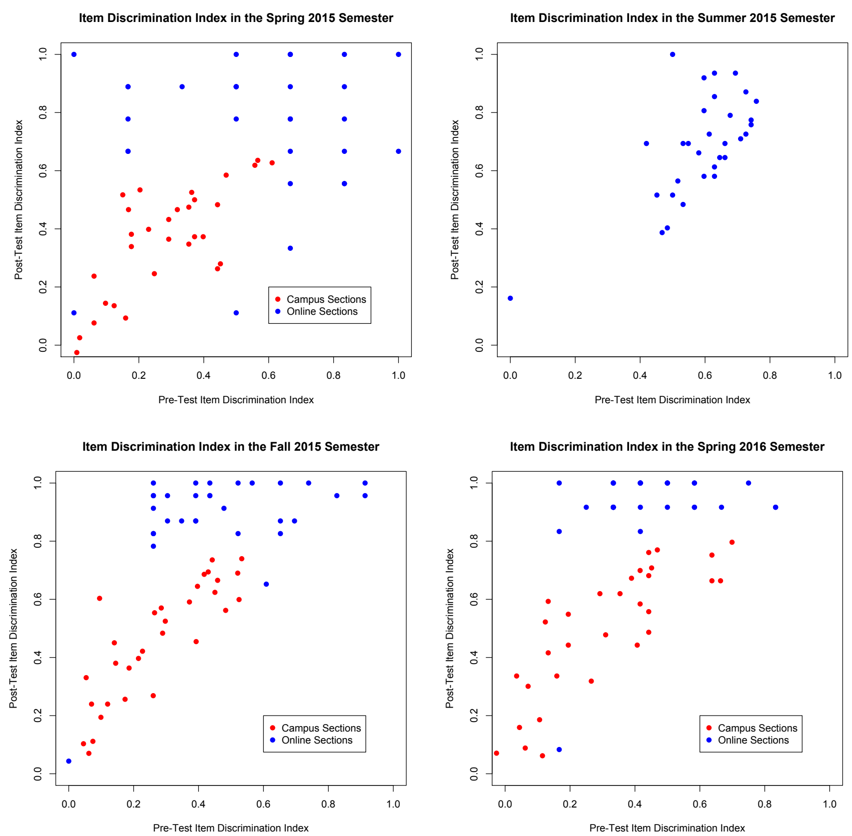

Figure A.4. The figure shows the item discrimination indices for the online and on-campus sections of the spring 2015 through spring 2016 semesters. Each point on the graph represents a question on the BEMA. In the spring 2015 graph, one of the questions had an index less than zero, so it does not appear on the graph. It is also important to note that the grid-like patterns of the online sections are an artifact of lower than ideal number statistics. Thus, they should not be interpreted as an underlying pattern in the data.

Point Biserial Correlation Coefficient

The point biserial correlation coefficient (r_{pbs}) measures the correlation between student's answers to an individual item on an assessment and their total score of the assessment. It is given by:

$$r_{pbs} = \frac{\bar{X}_c - \bar{X}_i}{\sigma_X}\sqrt{P(1-P)} \tag{A.3}$$

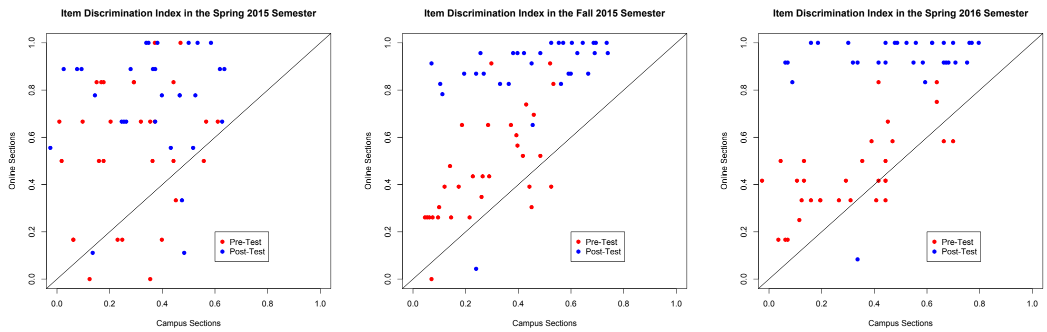

Figure A.5. The figure shows the pre-test and post-test item discrimination indices for the spring 2015 through spring 2016 semesters. Each point on the graph represents a question on the BEMA. In the spring 2015 graph, one of the questions had an index less than zero, so it does not appear on the graph. The lines with slope one are plotted for the sake of comparison.

In the equation above $\bar{X}_c$ is the average total score for students who correctly answer a given item, $\bar{X}_i$ is the average total score for students who incorrectly answered the item, σ_X is the standard deviation of the total score for all students tested, and P is the proportion of test takers who got the item correct [143].

The value of the point biserial correlation coefficient ranges between -1 and +1; a larger positive value indicates that a student's ability to answer a given item is predictive of scoring highly on the exam. Overall, the point biserial correlation coefficient gives a measure of how well an item discriminates students with higher and lower levels of knowledge. Ideally, one would like to see point biserial correlation coefficients that are as large as possible, but $r_{pbs} > 0.25$ is considered acceptable in many cases [144]

Figure A.6 compares the point biserial correlation coefficients between the online and on-campus sections. The graph on the left shows differences between online and campus students across all of the semesters. Online point biserial correlation coefficients for items generally exceed those seen for campus students. This means that students in the online section have a stronger correlation between their overall BEMA scores and their score on each question on the BEMA versus their on-campus colleagues. The graph on the right shows where the values tend to cluster for online and on-campus students and is consistent with the average correlations compared between the groups. Again, the online and on-campus sections are not experiencing the same test.

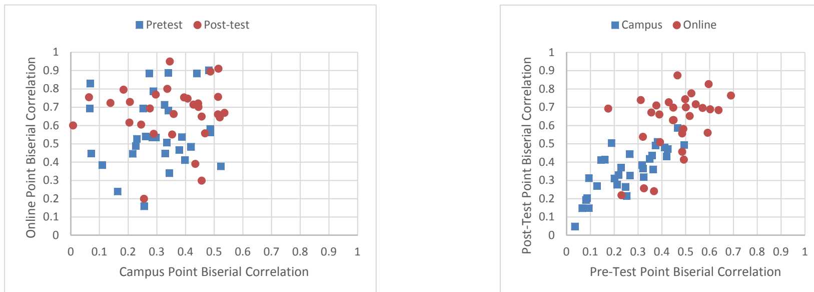

Figure A.6. The figure shows the point biserial correlation coefficient for the online and on-campus sections of the fall 2015 semester. The clustering of the data points shows that students are not experiencing the same test in the online and on-campus sections of Electricity and Optics.

Table A.6 gives the point biserial correlation coefficients in each section and each semester. Again, there is a distinct difference between how the different sections experience the BEMA concept inventory. We also see a huge difference between the pre-test and post-test coefficients in the online sections, suggesting that more students might be referencing external sources in the post-test than in the pre-test.

Ferguson's Delta

Ferguson's Delta is a statistic that measures the discriminatory power of an entire test by analyzing the distribution of total scores over the entire possible range of scores [145]. It is based on a comparison of the total scores between pairs of students. If a population of N students completes a given assessment, there are $N(N-1)/2$ pairs of students. The number of pairs of equal scores (P) is given by:

Table A.6
Average point biserial correlation coefficients for the BEMA pre-test and post-test in each semester. There is a distinct difference between the online and on-campus sections of Electricity and Optics.

Pre-Test Average Correlation			
Section	Spring 2015	Fall 2015	Spring 2016
Campus	0.364	0.345	0.352
Online	0.424	0.414	0.394

Post-Test Average Correlation			
Section	Spring 2015	Fall 2015	Spring 2016
Campus	0.448	0.431	0.484
Online	0.502	0.711	0.876

$$P = \sum_i \frac{f_i(f_i - 1)}{2} \tag{A.4}$$

where f_i is the frequency of each score in the sample. If we let K represent the number of items on the assessment, then the maximum number of unequal pairs (P_{max}) will occur when $f_i = N/(K+1)$. Ferguson's Delta (δ) is defined as the ratio of the number of unequal pairs of scores divided by P_{max}.

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - N^2/(K+1)} \tag{A.5}$$

The values for Ferguson's Delta range from zero to one where zero represents no well-defined distribution of scores and one represents a rectangular distribution of scores [146]. In Table A.7 we have calculated Ferguson's Delta for each section in each semester.

Table A.7
Ferguson's Delta for the BEMA pre-test and post-test for each section and semester.

Semester	**Campus (Pre)**	**Campus (Post)**	**Online (Pre)**	**Online (Post)**
Spring 2015	0.9497	0.9734	0.9132	0.9313
Summer 2015	NA	NA	0.9670	0.9516
Fall 2015	0.9492	0.9700	0.9456	0.8701
Spring 2016	0.9549	0.9736	0.8796	0.8001

Ideally, Ferguson's Delta should be greater than 0.90 to ensure that a test discriminates students with different skill levels appropriately [146]. This condition is met in all of the campus sections; however, the online sections in the fall of 2015 and spring of 2016 show values smaller than 0.90. This suggests that the discriminatory power of the BEMA concept inventory is not as high for the online sections.

Kuder-Richardson Reliability Index (KR-21)

The Kuder-Richardson Reliability Index (KR-21) measures the self-consistency of a test. If a test were administered to two groups of students at two different times, we would expect there to be a large correlation between the scores in each group, assuming that the students have approximately the same skill level and the test was administered in approximately the same conditions.

The correlation coefficient between the scores of two separate administrations of a given test is not necessarily a good way of determining the reliability index for many reasons. For example, students may remember test questions between examinations or the test conditions might differ significantly. Instead, we can take advantage of the fact that the BEMA concept inventory tests one specific knowledge domain - electromagnetism concepts. We can split this exam up into 31 "sub-tests" and compare how students perform on 31 parallel assessments. The KR-21 formula for the reliability index (r_{test}) is given by [147,148]:

$$r_{test} = \frac{K}{K-1}\left(1 - \frac{\sum P(1-P)}{\sigma^2}\right) \tag{A.6}$$

where K is the number of items on the assessment, P is the item difficulty of each question, and σ^2 is the variance of the total scores on the test.

An acceptable value for r_{test} depends on the purpose of the instrument. Generally speaking, if the index is higher than 0.7 it is acceptable for group measurements and if it is higher than 0.8 it is acceptable for individual measurements [146]. Table A.8 gives the values for the KR-21 index in each semester and section.

Table A.8
KR-21 reliability index for the BEMA pre-test and post-test for each section and semester.

Semester	**Campus (Pre)**	**Campus (Post)**	**Online (Pre)**	**Online (Post)**
Spring 2015	0.6492	0.7798	0.9368	0.9621
Summer 2015	NA	NA	0.9268	0.9388
Fall 2015	0.5831	0.7725	0.9064	0.9505
Spring 2016	0.6647	0.7816	0.9580	0.9706

First, we note that the KR-21 index is rather low for our on-campus sections, indicating that the reliability of the BEMA is not as high as we might like (Ding et al. reported a KR-21 index of 0.85 in their analysis [101]). However, what is particularly striking here is the huge difference between the online and on-campus sections. We are calculating suspiciously high KR-21 indices for the online sections of PHYS 24100 (greater than 0.90 in all cases). Again, the online and on-campus sections are not experiencing the BEMA exam in the same way.

A.5.2 External Validity

Data analysis, up to this point, has focused solely on the BEMA scores. Without an external measure of student ability to compare against we cannot unilaterally dismiss the possibility that the observed differences in the BEMA scores is due to differences in physics knowledge. The exams for this course have to be taken on campus by all students in the fall and spring semesters. During the summer semester, two midterm exams are administered online while one final exam must be taken on campus. This gives a standard way to measure student performance, and we can use performance on the exams to identify discrepancies in students' performance on the BEMA.

Importantly we are not claiming any rigorous equivalence between the course exams and the BEMA. We are instead interested in using the existing correlation between exam and BEMA performance to identify any anomalous performance exhibited by students.

Figure A.7 shows the total score on the first two exams versus the BEMA pre-test and post-test scores. The top row of graphs shows scores from the fall 2015 semester while the bottom row of graphs shows scores from the spring 2016 semester. We used the first two exams of the semester specifically because they covered the same topics that were on the BEMA. The final exam covered electromagnetic waves and optics, so it was omitted in this analysis.

The average item difficulty on our exams was much higher than the average item difficulty on the BEMA in each semester because we calibrated each test to have an average grade between 65 - 70% while it is not uncommon for students to score 50% or below correct on the BEMA. The graphs show that the vast majority of students in the on-campus sections (where cheating can be kept to a minimum) scored better on the exams than on the BEMA. However, many of the online students that were identified as cheaters due to the outlying data point (see Section A.6.2) are clustered in a specific region far to the right of the charts.

A.6 Discussion

A.6.1 General Trends

In each of the measurements above, we see that there is a large difference between the BEMA scores in the online and on-campus sections of PHYS 24100. The online section seems to outperform the on-campus sections on almost every question of the BEMA. Thus, we have to conclude that the BEMA is not a valid instrument for our online sections of PHYS 24100. This is almost certainly due to cheating on the BEMA (see Section A.6.2 below for details why).

Not everybody in the online sections of PHYS 24100 are cheating on the BEMA. Thus, the histograms of scores (Figure A.1) are made up of two different parts. First, there is a skewed bell curve that represents the students who attempt the BEMA in an honest way. These scores generally match up with the scores of their on-campus colleagues. Then, there is a group of outlying data points which correspond to perfect or near-perfect scores on the BEMA. These are the students that are cheating.

We find that it is simple to isolate those who are suspected of cheating on the BEMA exam. All of the individual item CTT measures from above are greatly skewed in the cases where students cheat on the BEMA. Additionally, the whole-test items are skewed when the online test environment is not as secure as originally thought.

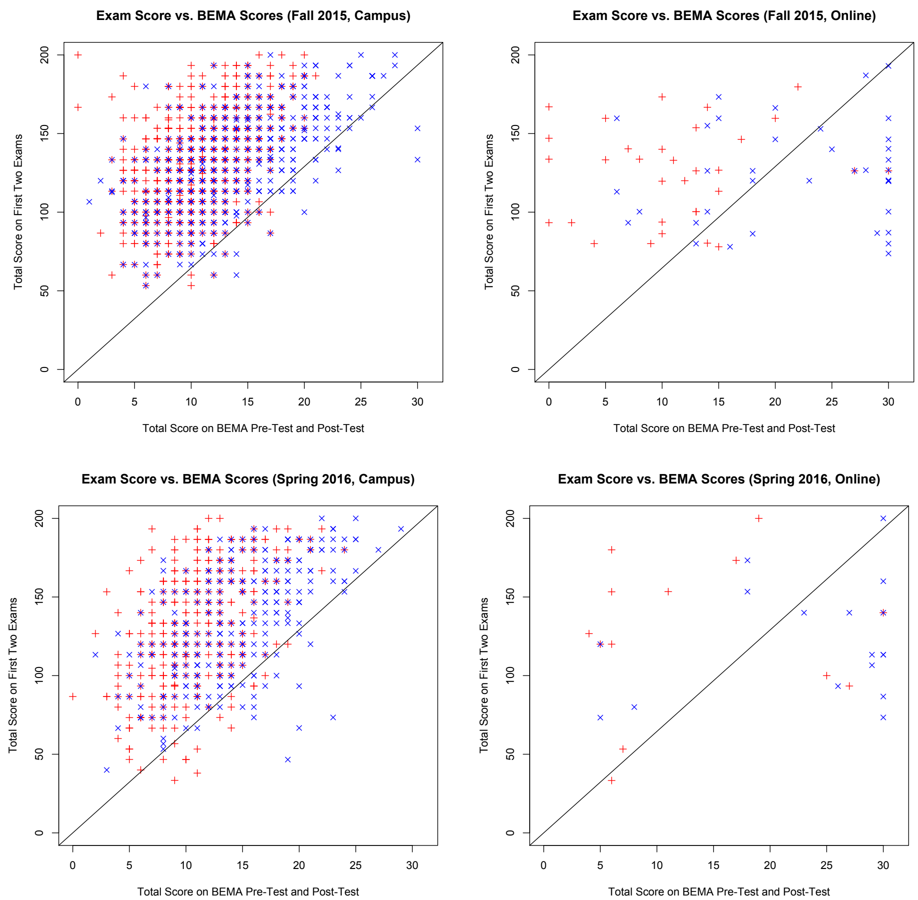

Figure A.7. A comparison of exam scores compared with the BEMA pre-test (red) and post-test (blue) scores in the fall 2015 and spring 2016 semesters. The fall 2015 semester is the top row of graphs while the spring 2016 semester is the bottom row of graphs. The graphs on the left show the data from the campus sections while the graphs on the right show the data from the online sections. Each point represents the total test score of one student on the first two exams (i.e. material that is covered on the BEMA). Thus, students who took both the BEMA pre-test and the post-test appear twice in each graph (i.e. one red data point and one blue data point). The online students who are identified as cheating are found on the right hand side of the graphs, corresponding to high BEMA scores but average exam scores.

Thus, the reliability and discrimination of the BEMA decrease significantly in an environment where many students are cheating.

We can also verify that students are cheating by conducting a test of the validity of the BEMA; the process is straightforward. First, administer the BEMA concept inventory as you would normally. Then, administer a short set of questions that cover the same topics as the BEMA, but have a lower item difficulty (call this the probe test). This can easily be done as a subset of a homework assignment or perhaps as a Clicker question in recitation. If you need to administer questions in an online environment, do not just pull questions from Google. Instead, reword questions so that they cannot be easily looked up. We used probe tests of 10-15 questions (i.e. our exams), but we predict smaller probes would work just as well. Finally, plot the standardized test question score versus the total BEMA score. Students who have a high likelihood for cheating will be clustered together in a region corresponding to high BEMA score, but low test score.

Although these methods might not be convenient for the average teacher in the average classroom, they are very easy to incorporate into a large, online homework system. Online homework systems can keep a running tab of the CTT measures from above, and identify any students who are moving into the realm of potentially cheating. Additionally, many homework systems like the ExpertTA [79] claim that they can not only change the numbers within a homework problem, but additionally reword the questions themselves. Systems like WebAssign [74] have built-in concept inventories (including the BEMA exam). It is a straightforward matter to find questions that are similar to the BEMA concept inventory and include them throughout the homework over the semester. Students that are tagged as high risk for cheating can be confronted on a case-by-case basis as per the wishes of the instructor.

A.6.2 The Outlier

Throughout the paper we repeatedly mention an outlying data point, and claim that it merits further discussion. Online sections outperformed the on-campus sections for almost every question on the BEMA - except for the outlying point. Our hypothesis was that this was due to cheating - online students were searching for BEMA solutions, and thus, outperforming their on-campus colleagues. However, it is extremely difficult to prove that high scores are, in fact, due to dishonesty rather than exceptional skill level.

Early in the project, members of the research group performed a series of Google searches in an attempt to see how easy it was to cheat on the BEMA; the results were not encouraging. It was very easy to find solutions to the BEMA and other concept inventories online. However, there was one serendipitous result that came from this search.

We found a website that posted the solutions to the BEMA questions near the top of the Google search results. However, the author of these solutions made one mistake on the list of answers. Thus, anybody who attempted to cheat from this list of answers would only get 30/31 questions correct. And to our amazement, the majority of the students suspected of cheating in our class did not have scores of 31/31. Instead, they had scores of 30/31 with exactly the same problem wrong between their tests and the online solutions.

We do not wish to ruin anybody's reputation by calling out a specific person. Thus, we provide two ways to access the slides without pointing the blame at anybody in particular. First, one can perform a Google search of the text of the first question on the BEMA: "The original magnitude of the force on the +Q charge was F...". As of the time we wrote this paper (August 2016), the web page was still available to the public. Additionally, if any reader would like to receive a PDF copy of the page we refer to, please contact us privately, and we will send you the page with the author's name blanked out.

This outlier is strong evidence that the online students were actually cheating on the test and not outperforming their on-campus colleagues. We want to stress that these answers were in no way planted or planned ahead of time. The research group did not realize that these answers were online when the BEMA was given to the classes.

A.7 Conclusions

We have shown that students in the online sections of our introductory electricity and magnetism class experience the BEMA differently than their on-campus counterparts. Thus, the BEMA is not a valid instrument for our online classes. The huge difference in the grade distributions between online and on-campus sections can be explained using a variety of CTT measurements. Normally, it is very difficult to determine if a student with a high score is cheating or is simply a skilled problem solver. However, in this case, we got very lucky; a thorough Google search for BEMA questions yielded a set of online answers to the BEMA that we believe students were using. This is because the BEMA scores in our online course from students who were suspected of cheating were consistently 30/31. It just so happened that the answers posted online had exactly one error, corresponding to the common mistake by all of our students.

Using this information, we determined that CTT measurements are skewed in very specific ways when students cheat on the BEMA. Additionally, a simple comparison of the BEMA pre-test and post-test results with results from in-class exams can identify cheating students based solely on a comparison of item difficulties alone. Although it would be difficult to implement these strategies in an individual classroom, they might find use when combined with a suite of other statistical techniques in an online homework system.

A.8 Acknowledgments

This work has been supported in part by the College of Science at Purdue University. We would also like to thank the PER@Purdue research group for their helpful discussions and insights.

B. TRANSCRIPTS

If any reader would like to receive copies of the focus group transcripts, please contact the author directly.

VITA

VITA

Cyrus Marazban Vandrevala, the son of Marazban and Khorshed Vandrevala, grew up in Libertyville, Illinois. After graduating from Libertyville High School, he earned a dual B.S. in biomedical engineering and physics from Marquette University in the December of 2009. From 2010 through 2016 he pursued a graduate degree at Purdue University, earning a Ph.D. in physics education research. Cyrus will be working at 8th Light after he graduates - a software consulting company based out of Chicago. During his free time, he hopes to combine his skills in software design and physics to design educational materials for physics students.