

How does a reformed course look after many years? A case study of the
reformed calculus-based introductory physics course at KSU.

by

Santosh Budhathoki

B.S., Tri-Chandra Campus, Nepal, 2009

M.S., Jawaharlal Nehru University, India, 2013

A THESIS

submitted in partial fulfillment of the
requirements for the degree

Master of Science

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
James T. Laverty

Copyright

© Santosh Budhathoki 2019

Abstract

We studied what happened after the reform, of the first semester, calculus-based introductory physics course, that took place almost 20 years ago. We examined how the physical environment, curriculum, the method of instruction, and the learning gain as measured by the FCI have changed over time since the reform.

The system that was put in place by the originators of the reform seems to be self-sustaining because the overall structure of the reform has remained largely intact. The course content and the lab experiments have remained almost the same. However, some minor modifications such as the introduction of the cooperative group problem-solving and the use of online homework are being used by some lead instructors teaching this course. The graduate students are introduced as primary instructors, LAs are being used as the secondary assistants in the studio, and clickers questions are being used in the lectures.

The FCI was used to assess the learning gain of the students and the success of the initial reform. The data show that the learning gain was higher immediately after the reform but has dropped more than a decade later. The focus of the research has been to explore the reasons for the fall in the learning gain in later semesters. Knowing why it dropped would give us information about how to sustain future reforms. We show that our data is reliable and use it to eliminate several possible explanations for the drop in the learning gain.

For future study, one may look at the amount of time faculty spend thinking about teaching and how that changes over time as a measure of “investment” in the course. One can also study how the group formation in the studio affects the learning gain. One may also take videos of the lectures and the studios and analyze them to see how is the distribution of time for the quiz/CGPS, lab experiment and the review of the homework problems in each studio, how much do the students actually get the opportunities to engage in active learning, how much do each students in a cooperative group contribute in learning. All these detail

video analyses can provide answers to the changes in the learning gain if there is any in the future.

Table of Contents

List of Figures	viii
List of Tables	x
Acknowledgements	x
Dedication	xii
1 Introduction	1
1.1 Concerns About Lectures	4
1.2 Course Reform	5
1.3 Improving and Assessing Student Learning	7
1.4 Sustainability	9
2 Context: The Initial Course Reform of the Calculus-Based Introductory Physics at KSU	11
3 Data and Analysis	16
3.1 Student Data	17
3.2 Physical Environment	19
3.3 Curricular materials	19
3.3.1 Lab Manual	19
3.3.2 Syllabi	19
3.4 Instruction	20
4 Results	21

4.1	Physical Environment	21
4.2	Curricular Materials	22
4.2.1	Textbooks Used	22
4.2.2	Exams	22
4.2.3	Use of Online Homework	23
4.2.4	Laboratory Experiments	23
4.2.5	Introduction of Cooperative Group Problem Solving (CGPS)	24
4.3	Instruction	25
4.3.1	Studio Primary	25
4.3.2	Number of Instructors	25
4.3.3	Learning Assistants	25
4.3.4	Lecture	26
4.3.5	Help-Room	26
4.4	Learning Outcomes	26
4.4.1	Withdrawal and Failure Rates	26
4.4.2	FCI Scores	27
5	Possible Explanations	32
5.1	Possible Reasons for the Fall in the Learning Gain	33
5.1.1	Hypothesis 1: Students in the earlier semesters were better prepared than those in the later semesters.	33
5.1.2	Hypothesis 2: Students who took the FCI are not representative of their classes	34
5.1.3	Hypothesis 3: Post-test scores are strongly influenced by when the FCI is taken	36
5.2	Can we explain the changes in the FCI scores with changes in the course?	38
5.2.1	Combination of homework type and quiz/CGPS in the studio	39
5.2.2	Changes in the studio instructors	42

5.2.3	Use of clickers in the lecture	45
6	Discussion, Conclusion, and Future Work	48
6.1	Limitation	49
6.2	Suggested Explanation	50
6.2.1	Newness Effect	50
6.2.2	Group Formation in the Studio	52
6.3	Recommendation for Future Research Work	53
6.4	Conclusion	54
	Bibliography	57

List of Figures

2.1	Experimental set up for measuring the acceleration of two-body system using Newton's law. Students did this before the reform.	14
4.1	Variation of the "DFW" rate with the semester before and after the reform of the course.	27
4.2	Variation of FCI pre-mean score with semesters. The error is based on the standard error of the mean for each semester. The FCI pre- mean scores vary from 0.46 to 0.49 for a period of almost two decades.	28
4.3	Variation of FCI Gain with semesters. The semesters with different colors are statistically significantly different. The error bars are the standard error of the mean. Semesters denoted by the same color with asterisk are also statistically significantly different from each other.	29
4.4	Variation of the Cohen's d of the FCI scores with the Semester.	30
4.5	The variation of the fractional gain for 8 different semesters taught by 5 different lead instructors.	31
5.1	Average Composite ACT score vs semester of students in the course. The error bars are the standard error of the mean of the composite score for each semester.	34

5.2	The bar graphs show the comparison of the fraction of students who got various grades (A,B,C,D, and F) in the course and took FCI to the fraction of the total students who got similar grades in the course for 6 semesters (A, B,C,D,F, and G). The data show that a higher percentage of students who got A grade in the course took the FCI test in the later semesters in comparison to the earlier semesters.	35
5.3	Variation of the Cohen's d of the various primary instructors for different studios in semesters E, F, G and H. Graduate students were working as primary instructors in semesters E, F, and H.	43
5.4	Variation of the gain of the various primary instructors for different studios in semesters E, F, G, and H. The error bars are the standard error of the mean. The variation of the gain of the studio sections with graduate students as primary instructors is similar to that of other groups and the gains not the lowest.	43
5.5	Variation of the fractional gain of the various primary instructors for different studios in semesters E, F, G, and H.	44

List of Tables

2.1	Weekly Instructional Sequence for the course. The first row is for the traditional course. The students do the laboratory work for two hours once a week and recitation for one hour, twice a week. The second row is for the reformed course. The students meet for two hours, twice a week in the studio.	15
3.1	Number of students who took the FCI in various semesters	18
4.1	Difference between paper and online homework	23
4.2	Difference between Quiz and Cooperative Group Problem Solving.	24
4.3	Data Summary.	28
4.4	Probabilities from ANOVA over gain across the semesters. The asterisk show statistical significance.	29
5.1	Data Summary of Week 12 VS Week 15 FCI results	38

Acknowledgments

I would like to express my sincere gratitude and appreciation to my supervisor, James T. (JT) Laverty, you have been a friend, an inspiration, a role model, a pillar of support, and a tremendous mentor for me. I would like to thank you for accepting me as a graduate student, encouraging and helping me every day, and guiding me in the successful completion of the project. Thank you for your affection and compassion.

I have great pleasure in acknowledging my gratitude to Sarah Peterson, my fellow researchers in the Physics Education Research Group at KSU, and Bryan Stanley. Your ideas, support, and encouragement have been great contributors to the completion of this project. Without Sarahs work and JTs guidance and supervision, this project would not have been completed and I shall be eternally grateful and indebted to them for their help, support, and guidance.

I would like to express my sincere appreciation to Dean Zollman, university distinguished professor and a pioneer of the Physics Education Research, and Alice D. Churukian, Teaching Associate Professor in Physics at the University of North Carolina for their valuable time, data and other resources, input and suggestions for my research work. I would also like to thank Tim Bolton, professor of physics at KSU, for providing valuable information about some of the changes that took place in engineering physics.

I would like to thank my supervisory committee members; JT Laverty, my lead supervisor and assistant professor in the department of physics, Michael J. OShea, professor and associate head of the department of physics, and Eleanor Sayre, associate professor in the department of physics, for serving as my committee despite their busy schedule. I would like to thank them for letting this defense be an online one.

I take great pride in acknowledging my sincere gratitude to Kim Coy, Peter Nelson, and Peggy Matthews in the department of physics for all the valuable materials they provided for this research work. I would like to thank Robert Auten and Roger Eaton in the registrar office

for all the data they provided and to Xianzhe Xue, information officer at the department of Statistics, for providing valuable input in the analysis of the data. This work would not have been completed without their support.

I would like to thank my fellow friends and compatriots for all their help, encouragement and support during my stay in Manhattan, Kansas. Last but not the least, I would like to express my sincere gratitude to my parents and sister for everything they did for me. Their love, encouragement, support, and blessings sustained me this far.

A special thanks to all the strangers in Manhattan who smiled at me and greeted me while walking to the department or back to my apartment. Thank you for being humble and kind. I will cherish the moment that I spent in the USA.

Dedication

I want to dedicate this thesis to my sister and my parents.

Thank you for your love and support.

Chapter 1

Introduction

Course reforms are complicated processes that require significant investment. Ideally, this investment is expected to provide not only short-term benefits but also long-term benefits to the students. Course reform requires a substantial amount of money, time, and organizing. The resources involved are

1. The monetary cost involved in the reform of the course

The traditional style of lecturing in a stadium-style lecture hall is probably the most cost-effective style of teaching. Switching to active learning system requires not only overhead costs such as building new infrastructure and purchasing new equipment but also long-term costs such as hiring more instructors and maintaining the system. For instance, in SCALE-UP classrooms for large classes, outfitting each group for a two-semester introductory sequence can cost \$4,000-5,000 per group (including tables, chairs, and computers)¹. However, this does not include the cost of renovating the space. Similarly, the research conducted at Florida International University confirms that the cost (of passing the introductory physics course) per student in an active learning system using modeling instruction is 30% higher than the traditional lecture system².

2. Time involved in the reform of the course

Reform of a course can't be done overnight. There is a lot of people and planning involved to make sure that the reform is achievable, sustainable and successful (in the sense that it meets its goals). For instance, the scale-up project at North Carolina State University took some years¹. Similarly the "New Studio" at our own Kansas State University took some years as well³.

3. The effort involved in the reform of the course

Often it may be difficult to convince the administration of the university to move from traditional lecture system to the active learning system because of the cost and other resources involved. Some faculty may be hesitant to be part of the reform because of the time they need to devote.

The financial cost, time, and effort involved in the reform may not be feasible for every university. For those courses that are reformed, ways of maintaining the reform successful for a long period of time need to be identified to ensure that the investment made in the reform provides fruitful results for a long period of time.

Physics Education Researchers have produced a lot of instructional strategies and curricular materials that help students learn better. These materials are disseminated through journals, workshops, seminars, conference, and presentations. They are valuable only when they are actually implemented in classrooms. Despite all these efforts, these materials are not incorporated significantly into the classrooms as anticipated and desired by the developers and the physics education researchers. Some of the reasons why the instructors continue to lecture despite the evidence that it is not effective are⁴

1. Faculty believe that the traditional methods are effective as it worked for them.
2. Faculty use the traditional methods because they are unaware of the research-based reforms.
3. Faculty are mostly interested in research and are not willing to spend time improving their teaching.

There have been significant efforts to disseminate the research-based instructional strategies by convincing the instructors to change their traditional style of teaching and adopting research-based practices. However, adopting such strategies at a secondary institution is not easy. For instance, the research done by Henderson and Dancy⁵ reveals that the five physics instructors who they studied faced varieties of situational challenges like students attitudes, expectations of content coverage, lack of instructor time, department norms, class size, etc that favored the use of traditional pedagogies. Also, there is potentially a mismatch between the expectations of the education researchers and the traditional faculties that may be a barrier for the spread of the reformed instructional practices.⁶

Although many physics instructors are aware of the problems associated with the traditional style of teaching and also know the ideas to improve by incorporating more research-based instructional strategies, the secondary implementations of these strategies are always challenging⁷. Moreover, the secondary implementation of these strategies in a new environment with different students, instructors and educational settings may not provide as much learning benefits as was observed with the original developers⁸.

Other work by Henderson and Dancy⁹ shows that even those instructors who had used at least one Research Based Instructional Strategies (RBIS), about 1/3 had discontinued using them. Some of the reasons for the discontinuation of the RBIS are the instructors faced students complaints¹⁰, the instructors were unable to cover the amount of course content that they feel was appropriate¹¹, and the results of the implementations were weaker than promised⁸. Hence, on one hand, the use of the RBIS is not as widespread as desired and on the other hand, there are cases of discontinuation, which is not desired at all. Now, another question arises, what about those courses that have continued using the RBIS? Have the results been as promising as they were before? In this thesis, we try to answer this question by studying a course that has been using RBIS for almost two decades. Specifically, we are interested in how the physical environment, curriculum, and instruction have changed in that time period; how the measures of success have changed, and what may have caused those changes.

1.1 Concerns About Lectures

Physicists have been trying to understand the struggles of the students in understanding the concepts of physics. The main goal has been to improve student learning. Traditionally, physics has been taught in stadium-style, large lecture halls in large universities in the USA. These lecture halls can accommodate hundreds of students simultaneously saving time and energy for their research works. Teaching so many students at a time saves space and money for the university as well. However, this style of teaching is not the most effective for a proper understanding of the materials taught.

In 1985, Halloun and Hestenes¹² published the result of their assessment of students basic knowledge before and after the instruction by designing two tests; a physics diagnostic test to assess students conceptual understanding of common physical phenomena and a mathematical diagnostic test to assess the mathematical skills of the students. These tests were given as pre and post-tests to nearly 1500 students taking university or college physics at Arizona state university and to nearly 80 students taking physics at a nearby high school. They concluded that a student's initial knowledge has a large impact on his/her performance in physics and conventional style of teaching produces little improvements in his/her basic knowledge. They also concluded that the basic knowledge gained under conventional instruction was essentially independent of the professor.

Other research shows that the traditional style of teaching physics has led many aspiring students to turn away from physics as they find the course dull, unwelcoming, and full of disjointed ideas.¹³ This style of teaching does not help students develop meaningful problem-solving skills¹⁴. Similarly, research shows that teaching science courses with lecturing and cookbook style laboratory experiments does not help students develop meaningful understanding of the materials taught¹⁵.

Many physics instructors following the traditional style of teaching spend time preparing standard lectures, giving explanations, demonstrating experiments during the lectures, and solving quantitative physics problems to illustrate problem-solving strategies. By doing so, they believe that the students are learning physics concepts and problem-solving strate-

gies, integrating them into a coherent set of ideas and are simultaneously developing higher critical thinking and reasoning skills. However, only a few students develop a functional understanding of the material taught and later become professional physicists. Research by Lillian Christie McDermott reveals that certain conceptual difficulties are not overcome by traditional instruction, a coherent conceptual framework is not typically an outcome of traditional instruction, growth in reasoning ability often does not result from traditional instruction, and teaching by telling is an ineffective mode of instruction for most students¹⁶.

These findings are robust as they have been demonstrated in many different institution types and with thousands of students. These issues are not because of the students or the instructors but of the method of instruction. Because of all these issues with the traditional style of teaching, there have been ongoing significant efforts to reform the introductory physics course.

Further research in physics education has shown that students can solve quantitative physics problems without having a deeper understanding of the core ideas needed to solve them¹⁷. Also, research in physics education and other STEM fields has shown that students learn and retain more when they are actively involved in the course, when they work together with their peers, and interact with their instructors rather than passively listen to the lecture¹⁸⁻²⁰. Hence the introductory physics course can provide better learning if it is student-centered rather than the instructor-centered.

1.2 Course Reform

All of these concerns about lecturing have led people to search for other methods of instruction that:

1. Increase student-instructor interactions.
2. Promote the active participation of students in the classroom.
3. Keep students engaged in learning.

4. Meet each student at their current level of understanding.

Over the last 4 decades, the research in physics education has led many physics departments to shift from traditional passive lecture system to active learning in many universities. Some of the goals of the reform of the introductory physics course are improving student understanding of core concepts of physics, increasing the attendance of the students in the class, decreasing the “DFW” rates (the percentage of students who withdraw from the course and those who receive “D” and “F” grades), developing student scientific reasoning skills, improving student attitude towards physics, and retaining the students in physics.

*Adapting to a changing world-challenges and opportunities in undergraduate physics education*²¹ states “An overarching theme has emerged from educational research: Learning improves when students are interactively engaged with their peers, their instructors, and the material being learned, and when they are integrating the newly learned concepts with their previous ideas, whether learned in a formal classroom or in everyday life. While this statement does not sound revolutionary, it does emphasize that success in learning is more strongly determined by how successfully and frequently students are engaged in the learning experience than by the content knowledge or the delivery skill of the instructor.”

“Active learning” (also known as “Interactive engagement”) has been found to be successful as the method is replicated in the physics department of universities other than where it was originally developed. There are clear evidence that the method is transferable as many departments in chemistry, biology, mathematics and some even in social sciences across the globe have adopted the active learning system.

Here we highlight a few active learning system relevant to our work.

1. Studio²²

Studio combines the lab and lecture part of the introductory physics course and makes students actively involved in learning the course materials via working together with their peers and interacting with the instructors. It was implemented for the first time at Rensselaer Polytechnic Institute in 1993. The class size for the studio physics is usually less than 50.

2. SCALE-UP (Students Centered Active Learning Environment with Upside-down Pedagogies)¹.

It is similar to the studio-format but it can accommodate about 100 students in a class. Much of the reading of the course material is done by the students at home, so students are already acquainted with the class materials. The lecture is not eliminated but it is substantially reduced and is used primarily to summarize and provide an overview of the topic. It was developed at North Carolina State University by Robert Beichner.

3. Peer Instruction¹⁷

In this, lectures consist of short presentations on key ideas and then a multiple-choice question related to the topic lectured is posed. Students think and record their individual answers and then try to convince their answers to the neighbors and make any correction to their previous answer. The lecturer gets the feedback from the students and then explains the correct answer and move to the next key concept. The lecturer repeats the above procedure again. There are pre-class reading assignments before each lecture. This was developed by professor Eric Mazur at Harvard University.

For a detail list of many other research based instructional strategies used in introductory physics, one may read Henderson and Dancy's paper²³.

1.3 Improving and Assessing Student Learning

A primary goal of course reform is to improve students' learning. A way to improve student learning is to reform the curriculum and change the method of instruction.

Assessing what students have learned is very fundamental in educational practice. Assessment helps us in knowing what the students have learned after the instruction. It aids us in understanding whether the reform is successful in achieving the desired learning goals or not. It also helps us in knowing which part of the curriculum is working well with the students. A commonly used method of assessment in research is the use of concept inventories. The most commonly used concept inventory is the FCI²⁴. It is a 30 item multiple-choice

questions designed to test student understanding of Newtonian concepts of force and motion against their common-sense beliefs or misconceptions. It is the most widely used concept inventory and there are national wide data available and these data show that the interactive engagement/active learning enhance student understanding of core concepts more than a traditional lecture-based system does^{18–20}. It is given as pre and post-tests, so it is an effective way of evaluating the effectiveness of the method of instruction and provides valuable information regarding the effectiveness of course reform. The other commonly used concept inventories in introductory physics are FMCE (Force and Motion Conceptual Evaluations)²⁵, BEMA (Brief Electricity and Magnetism Assessment)²⁶ and CSEM (Conceptual Surveys of Electricity and Magnetism)²⁷.

The assessment of the immediate results of the reforms using the concept inventories shows that the reforms have been successful. For instance, Robert Beichner and other researchers have compared data of over 16000 students over more than 5 years of instruction in traditional and SCALE-UP setting in several universities and found that in SCALE-UP setting, the conceptual understanding of the students increased, their attitudes towards the course improved, and failure rates drastically reduced¹. Similarly, Eric Mazur and Catherine Crouch have reported data from ten years of teaching with Peer Instruction in an introductory physics course at Harvard University. Their results show that students conceptual understanding as well as quantitative problem-solving improved upon implementing Peer Instruction²⁸. Similarly, using CLASP in an introductory physics course at UC Davies, the results in concept inventories were better and the attitudes of the students improved²⁹. At our own University (KSU), the implementation of the studio physics improved the conceptual understanding³⁰. All these successful results are from the immediate implementations of the reform.

The curriculum sets out what, how, and when a topic is to be taught. It supports all other parts of the system and guides day to day experiences in the classroom, the delivery of the content and the assessment of the student understanding. A well-structured and relevant curriculum if implemented effectively can provide better learning outcomes. For instance, several researchers have done a collective study³¹ of the measurements of the

performance of 2537 students in introductory electricity and magnetism courses at four institutions: Carnegie Mellon University, Georgia Institute of Technology, North Carolina State University, and Purdue University using two different curricula; a traditional curriculum and the matter and interactions (M & I) curriculum and have shown that the use of M & I curriculum can lead to significantly higher student post- instruction averages on the Brief Electricity and Magnetism Assessment (BEMA)²⁶ than courses using the traditional curriculum. The researchers conclude that M & I is more effective than the traditional curriculum at providing students with an understanding of the basic concepts of electromagnetism.

Although courses with active learning outperform nearly all traditional courses¹⁹, there is a wide variation in the normalized gain among the active learning courses. In Hake’s famous paper¹⁹, the normalized gain for the interactive-engagement courses vary from less than 0.20 to more than 0.64. However, this research was done not just with one course but with 62 introductory physics courses that implemented diverse active learning pedagogies. Hake put forward the hypothesis that the variations in the normalized gain was due to the course-to-course variations in the effectiveness of the pedagogy and/or implementations. Even with the secondary implementation of the same reform, Workshop Physics³² across multiple institutions, researchers have found the variation in the normalized gain from 0.39 to 0.57 as measured by the FCI³³. Similarly, the secondary implementation of the Tutorials in Introductory Physics³⁴ at a single institution in the same calculus-based introductory physics course but for various semesters, the variation of the normalized gain was found to be from 0.45 to 0.64 as measured by FMCE³⁵. These variations in the learning gain can pose a risk for the sustainability of the reform.

1.4 Sustainability

Research suggests, “Curricular and pedagogical reforms become institutionalized when organizational participants no longer perceive the reforms as special projects but as integral parts of organizational functioning³⁶. Further, the literature on educational change on STEM disciplines points to the structural support for the change process to be sustained. Kezar writes

“Change processes at the institutional level need some level of structural support over time to be fully institutionalized. On-going funding and operational support are needed, appropriate staffing, access to information, and other forms of support are necessary for a change to be sustained. Many changes have come and gone because they never had enough structural support, so they were the first to be removed in times of fiscal scarcity... For changes to be sustainable, they need to become part of the institutional structure, budgeting, and priorities³⁷. Literature also suggests that for the reform to persist, it needs to be valued and practiced by a larger group than the original innovators³⁸.

In the Physics department, the reform of the calculus-based introductory physics course has been institutionalized as it gets priorities and funding for its continuation. Also, the reform has been swiftly handed to other faculty members as the original developers no longer teach the course and is practiced by many faculty members and not just one, teaching the course. Even some modifications (such as the use of clickers) that have been introduced later, by a faculty, are being practiced by all faculty members teaching the course. Hence the faculty members are supportive of teaching and curricular reforms. These may be the reasons why the reform has sustained for almost two decades.

Although there has been a lot of work on the course reform, its effectiveness, and the dissemination of the reform efforts to other institutions, the long-term effects of the course reform and its sustainability remain an issue. In this thesis, we look at what happened after the secondary implementations with slight modifications of studio physics and focus on the long-term effects of the reform at a secondary institution. In this thesis, we look at the first semester of the calculus-based introductory physics course that was reformed almost two decades ago and see how it has evolved over time. We will especially focus on the following questions.

1. How have the physical environment, curriculum, and instruction changed over time after the reform?
2. How have the “DFW” rate and the assessment results changed over time?
3. How can we explain any changes in the learning gain that is observed in the data?

Chapter 2

Context: The Initial Course Reform of the Calculus-Based Introductory Physics at KSU

In order to investigate the long-term effects of course reforms, we present a case study of a reformed, calculus-based introductory physics course primarily taken by engineering and physics majors at Kansas State University, colloquially referred to as EP1. It is the first semester of a two-semester, sophomore-level course sequence. This course was reformed around the year 2000 and was the focus of Alice Churukian's dissertation³. In this chapter, we review that initial reform effort.

Before Spring 2000, EP1 followed a fairly traditional course structure of lecture, lab, and recitation. Faculty members delivered lectures in large lecture halls. The lecture served around 150 students and was delivered twice a week for one hour each. The students were divided into smaller groups of 40 students in recitation and 30 students in the lab. The recitation was taught by a faculty member for one hour, twice a week and the labs were taught by either a graduate student or a senior undergraduate for two hours, once a week. The labs were a single experiment with a lot of data taking. Each student then had a few days to write a complete lab report, which was several pages in length, and then turn it in.

The students had to do 13 labs throughout the semester and 12 of them were graded. The weekly schedule for this course can be found in the first row of Table 2.1.

In Fall 1994, the physics department examined the student evaluation of the course using a student evaluation tool, TEVAL, and also conducted individual interviews with students, faculty, and lab instructors. The department determined that the students were not satisfied with the course format (i.e. lecture/recitation/lab). Specifically, they had concerns about³⁰

1. The incompatibility of the lab and the lecture (i.e. the materials covered in the lab either lagged or led the materials covered in the lecture).
2. They found difficulties in making connections with the concepts learned in the lecture to the problems solved in the recitation and the experiments done in the lab.
3. The students noted with the written comments that the lab experiments were tedious, required to follow cookbook-style procedure with no creative thoughts, and only remotely related to the lectures.
4. However, the students preferred the smaller class size of the recitation and the lab. They also liked the increased number of student-instructor interactions in the lab and the recitations.

Based on this feedback, the department formed an ad hoc committee, responsible for proposing explicit changes. The committee met during the 1997-1998 academic year and developed the following goals to improve physics instruction at Kansas State University³

1. Improve the conceptual understanding of physics concepts for students who frequently focus on mathematical problem solving of physics and the number crunching in the laboratory.
2. Decrease the time which students spend capturing and manipulating data coupled with an increase in the time spent in the analysis of the data and related concepts.
3. Increase the time available for meaningful inquiry and discovery activities.
4. Increase the amount of professor-student and student peer instruction.

There existed major constraints for the change. The reform of the course could not significantly increase the teaching load of the research-oriented faculty members. Moreover,

any of the faculty members could be asked to teach the course, so the change had to be acceptable to all. Also, the physical space in the department was limited, so the change could not require a significant increase in the classroom space necessary to teach nearly 500 students enrolled each semester.

Because of these constraints, the committee decided to adopt a new format for the course, which they called “New Studio”. The New Studio was based on “Studio Physics” developed at Rensselaer Polytechnic Institute²². The New Studio format retained the lecture part of the course but combined the lab and recitation into a studio that can accommodate a group of 40 students at a time.

Two identical rooms were used for the studio session for the course. The rooms that were used as the labs were renovated for the studios. In a studio, there are 10 tables with a computer on each and 4 chairs around each. The computer contains the lab software and is connected to the internet. Also, near each table is a small blackboard where an instructor can discuss the problems with the students. At the front center, there is a large blackboard for the instructor and a projector.

The course content (syllabus) was retained. However, the format of the class was restructured and how the course was taught was redesigned. The New Studio format replaced the two-hours of lab and two hours of recitation per week with two, two-hour studios. The intention of combining the lab and the recitation in the same two-hour period in the studio was to successfully integrate problem-solving methods and conceptual skills with active hands-on activities and discovery-oriented lab activities. The studio was designed to have three parts: quiz, lab experiments, and review of the homework problems. The cookbook style labs were replaced by short, concept-oriented lab activities. For example, in a two-body accelerated motion experiment, before the reform, the students had to do pre-lab activity in which they were asked to derive the formula (this was done by the students at home) of the acceleration of the system of two bodies shown in Figure 2.1 and the formula was used in the experiment. All detail for each step of the lab and the table to be filled with the data were explicitly given. The students followed all those steps and measured the acceleration experimentally and compared it with the corresponding values predicted by Newton’s law.

However, after the reform, there were no explicit explanations of the steps involved and no table was provided. Instead, short questions that worked as hints were posed and students had to measure and compare the acceleration of the system. The goal was to make students think rather than just follow instructions.

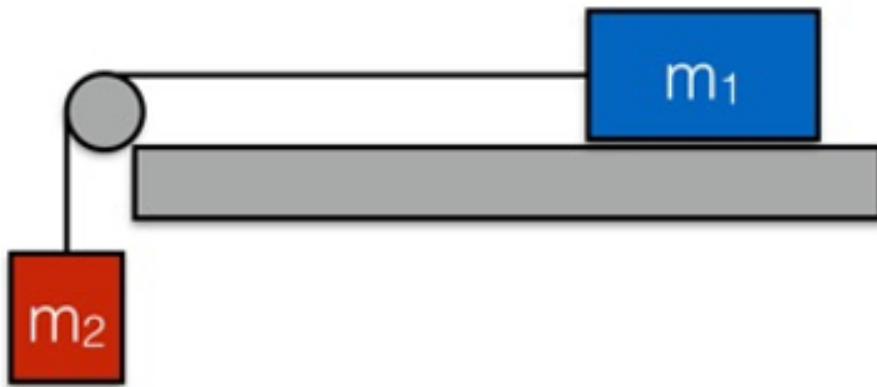


Figure 2.1: *Experimental set up for measuring the acceleration of two-body system using Newton's law. Students did this before the reform.*

The studio was taught by a team headed up by either a faculty or postdoc, known as the studio primary (or primary instructor), and a teaching assistant, also known as the studio secondary, who could be either a graduate student or a senior undergraduate. The primary is responsible for the quiz and the review of the homework problems whereas the secondary is responsible for the laboratory part of the studio. Both the primary and the secondary move around the studio to assist students for most of the class.

The students are first taught the physics concepts and problem-solving skills on Monday's and Wednesday's lecture and are asked to implement them on Tuesday's or Wednesday's studio and on Thursday's or Friday's studio. Hence, the course is well coordinated now as the materials covered in the lectures lead the materials covered in the studio. The weekly schedule for the reformed course can be found in the second row of Table 2.1.

The students randomly choose the table at the beginning of the semester. After every exam, they move to a different table.

	Monday	Tuesday	Wednesday	Thursday	Friday
Traditional Course	Exams at 3 weeks interval	Lecture, Lab	Recitation, Lab	Lecture, Lab	Recitation, Lab
Reformed Course	Lecture, Instructional Meeting	Studio all day	Morning Studio, Lecture in afternoon	Studio all day	Morning Studio, Exams at 3 weeks interval

Table 2.1: *Weekly Instructional Sequence for the course. The first row is for the traditional course. The students do the laboratory work for two hours once a week and recitation for one hour, twice a week. The second row is for the reformed course. The students meet for two hours, twice a week in the studio.*

On a typical day in the studio, students are expected to devote half of the time to quiz and homework discussion and half of the time in performing the lab experiments. The students are expected to go through the lecture material, solve homework problems, and read relevant portions of the studio manual for the next day's studio. The students take the quiz independently at the beginning of the studio. The studio primary then solves the quiz on the board. The studio secondary demonstrates the lab activities and the students do them in groups. They do about 28 lab activities in a semester and 24 of them are counted for the final grade. Each student writes the lab within the 2 hours period of the studio but only one out of the four in a table is graded and everyone in that table gets the same score. The goal is to encourage group discussion and enhance peer instruction. In some days, in a studio, they may do two lab activities as well. Often a few homework problems are discussed if the students have problems and if there is enough time left.

It has been nearly two decades since the calculus-based introductory physics course was reformed and the initial research done by Alice Churukian. She concluded that the reform was successful in improving the learning gain as measured by the FCI.

In this thesis, I am looking at how the course reform (physical environment, curriculum and the instructors/method of instruction) has changed over time, how the learning gain as measured by the FCI has changed since then and try to explain the observed changes.

Chapter 3

Data and Analysis

Henderson and fellow researchers have developed four broad categories of change strategies in Undergraduate STEM courses³⁹. The initial change made in this course falls in category 1 (Disseminating: CURRICULUM AND PEDAGOGY). As discussed in chapter 2, the department identified the problems associated with the course and also knew that the problems were not unique to KSU but were in agreement with those at similar institutions nationwide³⁰. The department also might have some ideas of how other institutions tried solving these problems and developed the vision to change the course by designing new environment, curriculum, and changing the method of instruction.

As discussed by Borrego and Henderson, STEM undergraduate education researchers usually focus on disseminating curriculum and pedagogy as the change strategy. They write “In fact, discussion about how to improve undergraduate STEM instruction is typically conceptualized solely within this category⁴⁰. At KSU, the original developers adopted some ideas of studio physics developed at Rensselaer Polytechnic Institute and modified the physical environment, curriculum and the method of instruction within the available resources and the constraints of the department. Hence, to address our research questions, we collected data on the physical environment, curriculum, and instruction/instructors since the reform of the course. The data collection took place near the end of the project, thus restricting us from collecting all possible information that could possibly provide more insights into the

project.

3.1 Student Data

We collected FCI data from eight different semesters (which we will call semesters A-H), one before the implementation of the studio and seven after the implementation of the studio. The semesters B, C and D represent consecutive semesters. Semester E is nearly a decade later than semester D and semester F is about six years later than semester E. Semesters F, G, and H represent consecutive semesters. These gaps account for a timespan of nearly two decades over all eight semesters.

The FCI was given as a pre and post-test in each semester. Before the implementation of the studio, the test was given in the laboratory. In all subsequent semesters, the data was collected as part of the studio. The pre-test was given at the beginning of the semester and the post-test was given either three quarters of the way through the semester (week 12) or at the end of the semester (week 15, this will be discussed in more detail in chapter 4). In all semesters, the students were awarded points equivalent to that day's studio just for taking the FCI (i.e. participation credit). The FCI for the first four semesters was given to measure the effectiveness of the original reform. In semester E, the FCI was given to see if the scores remained similar to the results just after the reform, but no thorough analysis was done at that time. In semesters F, G, and H, it was given to measure the effectiveness of a new LA Program⁴¹. The Table 3.1 shows the number of students who took FCI pre-test, post-test and the number of matched data points for the 8 semesters. However, in all these semesters, we used only the matched data.

We ran an ANOVA on the pre-test scores across semesters to ensure that we can run ANOVA on gain across semesters for the subsequent analysis.

We also ran an ANOVA with Bonferroni correction on the gain across semesters. The gain for any individual student is defined as

$$Gain = (PostScore\%) - (PreScore\%) \quad (3.1)$$

Semester	Pre-Test	Post-Test	Matched Data
A	107	100	99
B	135	128	116
C	275	244	201
D	119	101	89
E	307	254	237
F	448	378	332
G	254	259	128
H	314	315	313

Table 3.1: *Number of students who took the FCI in various semesters*

We also plotted Cohen’s d vs. semester and FCI fractional gain vs. semester. Cohen’s d is defined as follows:

$$Cohen's\ d = ((PostMean\%) - (PreMean\%))/sd \quad (3.2)$$

Cohen’s d is a measure of how important the change made is. It is a measure of how far the mean has shifted relative to the standard deviation. The suggested values for the small, medium and large Cohen’s d are 0.2, 0.5 and 0.8. Large value of Cohen’s d means the change made is important and the small value of Cohen’s d means the change made is unimportant. The standard deviation (sd) is the pooled standard deviation of the pre- and post- test scores.

The FCI fractional gain is defined as follows:

$$Fractional\ gain = (PostMean - PreMean)/(100\% - PreMean) \quad (3.3)$$

We also obtained students’ course grades and their incoming ACT scores from the registrar’s office for all students enrolled in the course over the nearly 20- year period we are examining. We analyzed the incoming students’ composite ACT scores using linear regression.

We first calculated how many students got different grades in the course in each semester. We then generated a linear regression for “DFW” grades over the semesters and used Root Mean Square Error (RMSE) to determine the model for best linear fit. We also matched

overall students' grade in the course with the grade of those students who took FCI by matching their IDs (we were able to match the IDs for semesters A, B, C, D, F and G as the IDs were same). Then we generated bar graphs and compared how the various grades of students who took FCI vary with the corresponding grades of the overall students in the course for different semesters.

3.2 Physical Environment

We collected data such as where the lecture, exam, studio classes took place for all the semesters since the reform of the course. The data were collected from the physics department's website and the analysis of the data involves keeping record of any changes that is seen in rooms used or remodeling of the spaces.

3.3 Curricular materials

3.3.1 Lab Manual

We do have the lab manual for two semesters, one almost immediately after the reform and the other a decade later. The later one is the most recent lab manual used for the course. For the analysis, we compared the titles of the lab and the list of the experiments in the two lab manuals.

3.3.2 Syllabi

We collected the course syllabi of all the semesters from Spring 1999 to Fall 2001 and Fall 2008 to Spring 2018. From each syllabus we extracted the following information:

1. Textbook used.
2. The number of exams (semester exams and the final exam).
3. Whether the semester exams were cumulative or not.

4. Whether there was a quiz or cooperative group problem solving (to be discussed later in subsection 4.2.5) in the studio
5. Total points allocated for the studio including the homework.
6. Total points allocated for the exams.
7. Whether there was online homework or paper/written homework.
8. Whether the clickers questions were used in the lectures or not.
9. When were the materials asked in the FCI covered in the course?

For the analysis, we made a table of all the variables mentioned above and compared the changes, if any, in those variables for the semesters with FCI data.

3.4 Instruction

We collected information on studio primary instructors for all the 8 semesters with FCI data. The list was collected from the departmental websites. From this information we focused on the number of instructors in a studio (primaries and secondaries) and the academic position of the studio primary and how those changed over time.

Chapter 4

Results

In this chapter, we look at how the physical environment, curricular materials and the method of instruction have changed since the initial reform of this course. We also look at how the learning gain as measured by the FCI and students' failure rate as measured by DFW changed over the semesters.

As mentioned in Chapter 3, we do have 8 semesters of FCI data. These 8 semesters were taught by 5 different lead instructors who we name as Professor Cross, Professor Square, Professor Star, Professor Circle, and Professor Triangle.

4.1 Physical Environment

The lecture takes place in the lecture halls (labeled as small, medium, and large). The large hall can accommodate 348 students, the medium can accommodate 176 and the small hall can accommodate 112 students. The medium hall has been used as the lecture hall since before the reform of the course. In spring 2009 and spring 2010, both the small and the medium halls were used as the lecture hall. Both the small and the medium halls are not that big in comparison to the large hall, which has been primarily used as the room for the exam since before the reform. However, the large hall was renovated and since spring 2017 it has been used as the lecture hall besides the medium hall. The small and the medium halls

contain fixed chairs but the large hall contains chairs that are bolted but can be rotated. The exams take place in all three halls (if the large hall cannot accommodate all the students for the exam, then the other halls are used). Since Fall 2014, two studios have been used for the first semester of the course.

4.2 Curricular Materials

We analyzed both the lab manual and syllabi looking at changes in the following variables over the semesters.

4.2.1 Textbooks Used

The course continued to use *Fundamentals of Physics* by Halliday, Resnick, and Walker for 8 years after the reform. Then one of the lead instructors (Professor Circle) started using *Physics for Scientists and Engineers* by Tipler and Mosca and has been using the same book since then. The same book (*Physics for Scientists and Engineers*) was used for the three consecutive semesters since its introduction. After that, the two textbooks were used almost alternately by different lead instructors except for one semester when *University Physics* by Young and Freedman was used and the lead instructor (Professor Triangle) who used this later reverted to *Fundamental of Physics*, which he was using before. For all the semesters with FCI data, the textbooks used were either *Fundamentals of Physics* or *Physics for Scientists and Engineers*.

4.2.2 Exams

Since the reform, students take five mid-term exams and a final. Each mid-term exam is worth 10% of the total grade points and the final is worth 20%. The mid-term exams occur at regular intervals (almost every 2-3 weeks) and none of these exams are cumulative. The lowest score obtained in the mid-term exams is dropped. Some lead instructors replace the dropped score by the average of the score obtained in other four exams to make the total

possible points 50% while some professors completely drop it making the total possible points to 40%. The final exam is cumulative and compulsory to pass the course. Depending on the faculty member teaching the course, the total points for the exams vary from 60% to 70% points.

4.2.3 Use of Online Homework

Since fall 2008, Professor Circle has been using online homework using WebAssign. In another semester, Professor Triangle used Mastering Physics for online homework for one semester near three quarters since the reform. So far, five different lead instructors have adopted using online homework in fourteen different semesters. However, Professor Triangle still uses paper/written homework. Most importantly, for both types of homework, students solve them twice a week. The Table 4.1 shows the difference of the paper and online homework for the course we are focused on.

	Paper Homework	Online Homework
Graded By	Teaching Assistants	Online homework system
Graded Problems	1 per assignment	All
Retries on missed problems	None	Usually 5 retries on each problem
Feedback	TA can give specific feedback on the problem that is graded	Feedback given when the problem is missed for some of the problems
Partial Credit?	Yes	No

Table 4.1: *Difference between paper and online homework*

4.2.4 Laboratory Experiments

The lab manual was revised in 2004 and 2005 to add some problems at the end of each lab. We don't know if people used them at some point, but we know they haven't been used in the later years. The lab experiments are being pulled from the same set of experiments and anecdotal evidence tells us that the students are doing almost the same labs.

With the passage of time, the old lab equipment has been replaced by more updated versions. For example, there is no air track anymore. It is replaced by dynamic tracks for the study of linear and accelerated motion, momentum and collision, Newton's second law etc.

4.2.5 Introduction of Cooperative Group Problem Solving (CGPS)

The studio consists of basic three components; review of the homework problems, lab experiments, and the quiz/CGPS. Since fall 2008, Professor Circle has replaced the quiz by the cooperative group problem solving (CGPS)^{42, 43}. In CGPS, the students work together in a group of four to solve an open-ended physics problem. The instructors neither provide solutions to the problem nor model the problem but observe students working in groups and provide help as well as feedback as needed. The CGPS has been used by three other lead instructors teaching this course. Since its introduction in fall 2008, four different lead instructors have implemented it in 14 different semesters so far. However, Professor Triangle still uses the quiz. The following Table 4.2 shows the difference between quiz and CGPS, here at KSU.

	Quiz	Cooperative Group Problem Solving
Any help from the peers or the instructors?	No, students do independently	Yes, students often ask instructors for help whenever they need
Time required	15-20 minutes	about an hour (because of longer time required, there is less or no time for homework discussion)
Who makes the problems?	The primary instructor	The lead instructor
Uniformity across all studios?	No	Yes

Table 4.2: *Difference between Quiz and Cooperative Group Problem Solving.*

4.3 Instruction

In the time frame of the study, there have been notable changes in instruction, including the introduction of graduate students as studio primaries, changes in the number of instructors, introduction of learning assistants (LA), use of clickers in the lecture hall and the extension of the help-room hours. We discuss each of these changes in more detail below.

4.3.1 Studio Primary

In semesters immediately after the reform, the Studio Primary was either a faculty member or a postdoc. For instance, in spring 2000 and spring 2001, there were 4 studios and all of them were taught by faculty members. In fall 2000, there were 6 studios with faculty members as primary instructors and 2 studios with postdocs as primary instructors. In later semesters, graduate students were assigned as the primary instructors and some non-tenure track instructors were hired by the department to work as the primary instructors.

4.3.2 Number of Instructors

As discussed in chapter 2, each studio is taught by the two instructors. However, a couple of semesters after the reform, the department experimented with 3 instructors in a studio; a faculty, a teaching assistant and the third one known as the class assistant. The class assistant was assigned to assist the students while they were working on the hands-on activities in the studio and was not given any responsibilities, such as grading the studio's work, outside of the studio. The department later determined that one of the instructors in the studio was idle, so reverted to two instructors in a studio soon after.

4.3.3 Learning Assistants

Learning Assistants (LAs)⁴¹ are undergraduate students who took the same course recently, did well in the course, and were recruited to work as instructors in the studio. For this course,

the LA works similarly to a teaching assistant (TA) but the main difference between the two are that LAs take a weekly pedagogy course taught by a PER faculty member whereas TAs do not take. Also, there are two LAs in a studio but only one TA in a studio. In spring 2017, the physics department introduced LA program and has been used since then in the studio.

4.3.4 Lecture

In Fall 2008, Professor Circle started using clickers in the lecture hall and since then students have been using clickers in the lecture halls to give responses to short multiple-choice questions posed intermittently by the lecturer.

4.3.5 Help-Room

Usually, one or two graduate teaching assistants or senior undergraduate teaching assistants are assigned to help any incoming students, especially in their homework problems. In later semesters, the help-room hours were extended. In Fall 2000, the help-room hours were 8 hours per week but in later semesters, it is extended to 30 hours per week.

4.4 Learning Outcomes

We do have withdrawal and failure rates as measured by the DFW rate and the learning gain of the students as measured by the FCI data (Cohen's d and the fractional gain, which are defined in chapter 3).

4.4.1 Withdrawal and Failure Rates

The withdrawal rate is measured by the percentage of the students who withdrew from the course and the failure rate is measured by the percentage of the students who got D and F grades in the course. We have combined them together and Figure 4.1 shows the “DFW” rate of students in percentage before and after the reform of the course. The graph shows

that the DFW rate was higher but decreasing before the reform. After the reform, it has leveled out at about the same level it had reached before the reform. However, variability has increased. The values of Root Mean Square Error (RMSE) for 2-lines model is 6.26. If we use a single linear model for the entire DFW data, the value of RMSE increases to 7.09, so two separate linear models is a better fit than a single fit to our DFW data.

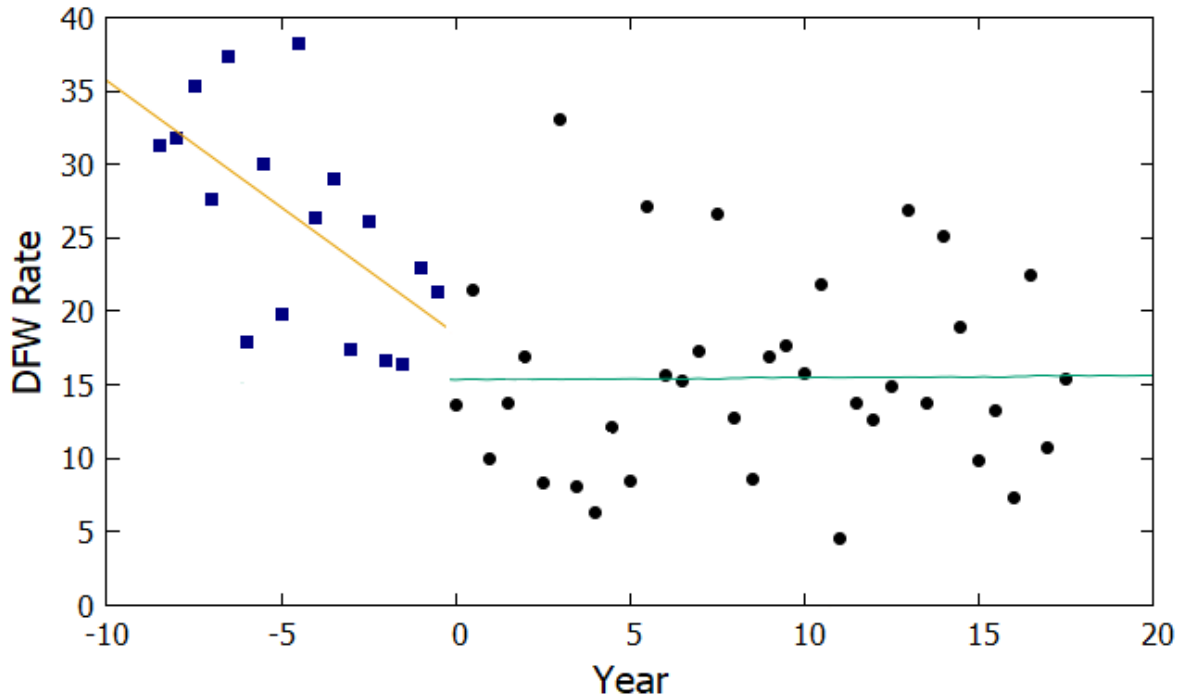


Figure 4.1: *Variation of the "DFW" rate with the semester before and after the reform of the course.*

4.4.2 FCI Scores

As mentioned earlier, we have 8 semesters of FCI data, clustered primarily at the time of the initial reform and at the end of our analysis period. We want to know if the FCI gains for different semesters are statistically significantly different or not. Our pre-test scores were statistically the same, as shown by ANOVA test and are plotted in Figure 4.2, so this allowed us to compare FCI gain using ANOVA test with Bonferroni correction to see if the FCI gain for the 8 semesters were statistically different from one another.

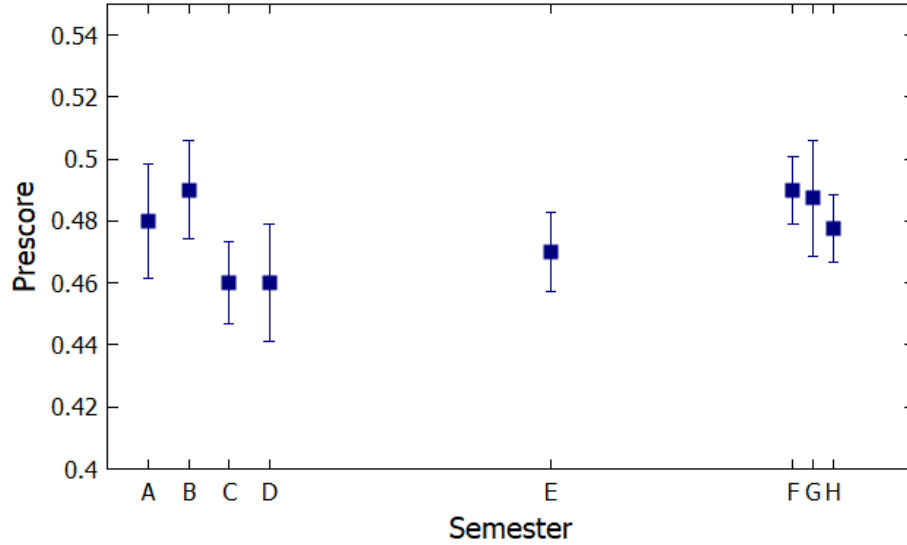


Figure 4.2: Variation of FCI pre-mean score with semesters. The error is based on the standard error of the mean for each semester. The FCI pre- mean scores vary from 0.46 to 0.49 for a period of almost two decades.

Table 4.3 shows the FCI pre-test mean scores, Post-test mean scores, Standard deviation (SD), Cohen's d and the fractional gain for the 8 semesters with the FCI data.

Semesters	Pre-Test Mean	Post-Test Mean	SD	Cohen's d	Fractional Gain
A	0.48	0.57	0.32	0.47	0.17
B	0.49	0.70	0.29	1.53	0.41
C	0.46	0.68	0.25	1.67	0.41
D	0.46	0.67	0.25	1.61	0.39
E	0.47	0.57	0.29	0.69	0.19
F	0.49	0.63	0.13	1.08	0.27
G	0.48	0.60	0.29	0.86	0.23
H	0.48	0.60	0.26	0.97	0.23

Table 4.3: Data Summary.

The plot of the FCI Gain VS Semester is shown in Figure 4.3 and probabilities of ANOVA of gain across all semesters are shown in Table 4.4.

We also plotted the Cohen's d and the fractional gain (for historical reasons) of the FCI data. In social sciences, researchers use effect size whereas in PER the researchers have historically used fractional gain since it was introduced by Hake¹⁹. Both the effect size and fractional gain are measures of the learning gain. Figure 4.4 show the variation of the

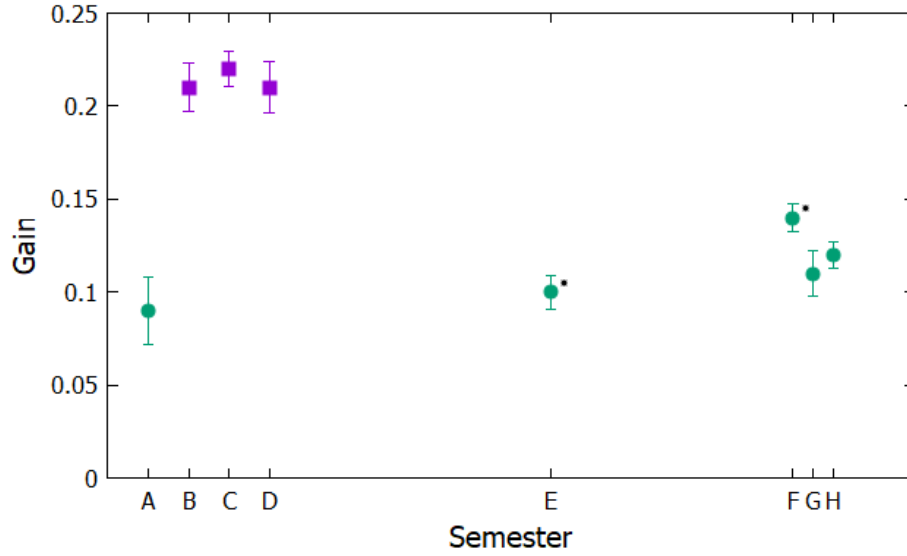


Figure 4.3: Variation of FCI Gain with semesters. The semesters with different colors are statistically significantly different. The error bars are the standard error of the mean. Semesters denoted by the same color with asterisk are also statistically significantly different from each other.

Cohen's d for the 8 different semesters with FCI data.

Figure 4.5 shows the variation of the fractional gain for the 8 different semesters with FCI data.

Semesters	A	B	C	D	E	F	G	H
A		<0.0001 *	<0.0001 *	<0.0001 *	1.0000	0.0639	1.0000	1.0000
B			1.0000	1.0000	<0.0001 *	<0.0001 *	<0.0001 *	<0.0001 *
C				1.0000	<0.0001 *	<0.0001 *	<0.0001 *	<0.0001 *
D					<0.0001 *	0.0054*	<0.0001 *	<0.0001 *
E						<0.0001 *	1.0000	0.4170
F							0.2248	0.2534
G								1.0000
H								

Table 4.4: Probabilities from ANOVA over gain across the semesters. The asterisk show statistical significance.

The ANOVA on the FCI gain versus semester revealed the following:

1. Semesters B, C, D are not statistically significantly different from one another but they are statistically significantly different from other semesters.

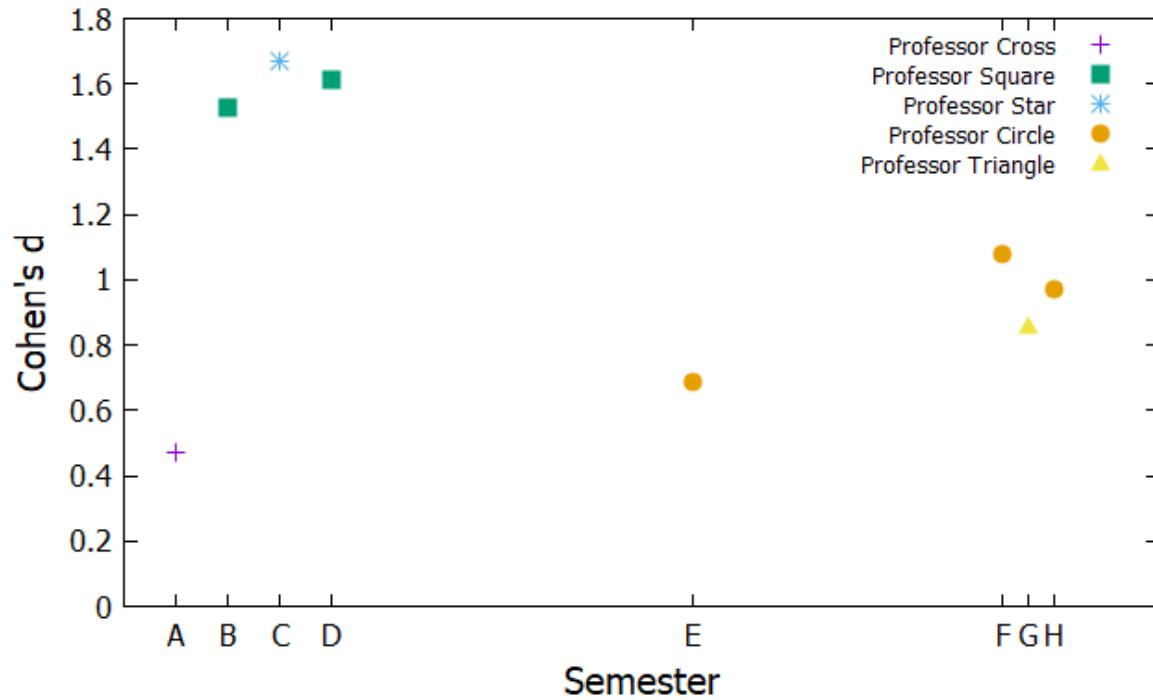


Figure 4.4: *Variation of the Cohen's d of the FCI scores with the Semester.*

2. Among semesters A, E, F, G and H, only semesters E and F are statistically significantly different although both of them are taught by the same lead instructor with the same curriculum and method of instruction.

More details of the results can be found in Table 4.4.

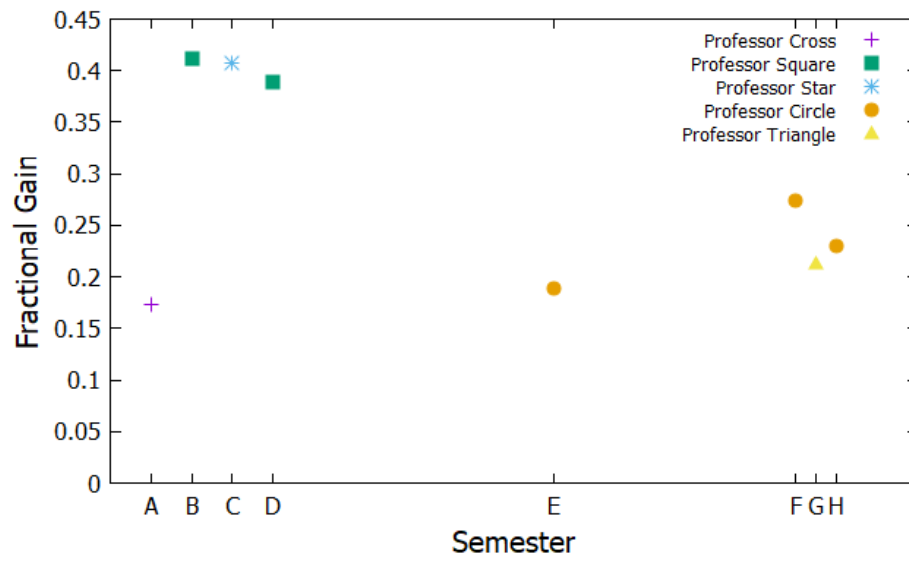


Figure 4.5: *The variation of the fractional gain for 8 different semesters taught by 5 different lead instructors.*

Chapter 5

Possible Explanations

Figures 4.4 and 4.5 show the drop in the Cohen's d and the FCI fractional gain for later semesters in comparison to the semesters immediately after the reform of the course. According to Hake's criteria¹⁹, the fractional gain for later semesters are low (less than 0.3), comparable to that of the traditional course. As mentioned before, considering the time and resources that go into the reform of a course, the goal would be for the effects of the reform to be lasting over a long period of time. Having discussed how the course structure, curriculum, instructors/method of instruction, DFW rate, and the learning gain as measured by FCI changed with time since the reform, we try to dig deeper and see if we can find the reason for the fall in the gain.

Before we try to dig deeper and try to find the reasons for the fall in the FCI gain, we need to answer another question, which is, “can we trust the data that we have?” The drop in the fractional gain can be related to something other than the course structure, curriculum, and the method of instruction. The following is a list of possible hypotheses, that we will explore, that could lead to the observed lower gains in later semesters.

Hypothesis 1: Students in the earlier semesters were better prepared than those in the later semesters.

Hypothesis 2: Students who took the FCI are not representative of their classes.

Hypothesis 3: Post-test scores are strongly influenced by when (12th or 15th week) the FCI is taken.

After discussing these hypotheses, we will try to answer if the drop in the fractional gain is due to the changes in the course after the reform.

5.1 Possible Reasons for the Fall in the Learning Gain

Here we discuss the hypotheses that could possibly lower the FCI fractional gain in the later semesters

5.1.1 Hypothesis 1: Students in the earlier semesters were better prepared than those in the later semesters.

The equation for the gain explicitly involves the pre-test scores which is related to the students' preparation before they join the introductory physics course at KSU. The students' preparation can have an impact on the normalized gain as defined by Hake¹⁹. If the students are coming with better preparation, they may do better in the course throughout the semester. A study done by Coletta and Philips⁴⁴ found a positive correlation between student's SAT scores and the normalized gain. However, a recent secondary analysis on published data about FCI and FMCE scores and gains shows that the fractional gain did not correlate with the SAT scores between 25th and 75th percentile of the incoming students²⁰.

We are using the ACT scores and FCI pre-scores as measures of the preparation of the students coming into the course. If the students in earlier semesters were better prepared, the less prepared students in later semesters may be the reason for the drop in the FCI gain.

The Figure 4.2 shows the variation of the FCI pre-mean score in fraction over the semesters. The Figure 5.1 shows the variation of the ACT composite mean with the semesters. From Figure 4.2 and Figure 5.1, it seems that the students in later semesters are equally well prepared as measured by the FCI pre-mean scores and slightly better pre-

pared as measured by the ACT scores. This suggests that the drop in FCI gain may not be the result of changes in how well-prepared students are in different semesters.

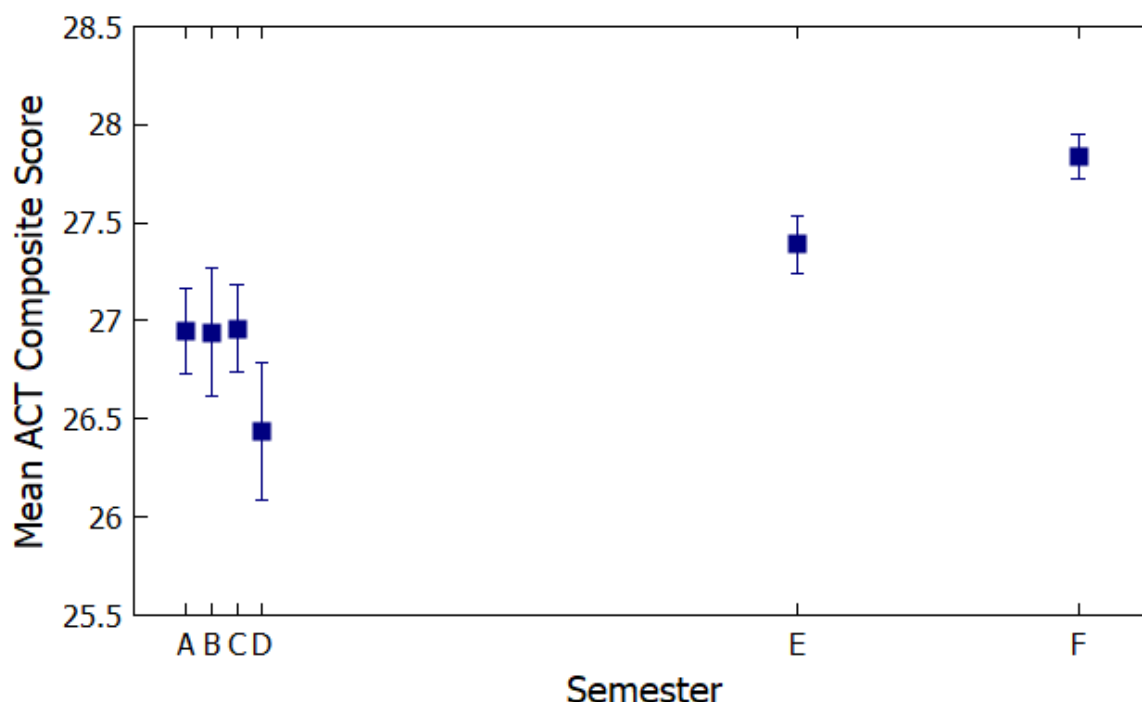


Figure 5.1: *Average Composite ACT score vs semester of students in the course. The error bars are the standard error of the mean of the composite score for each semester.*

5.1.2 Hypothesis 2: Students who took the FCI are not representative of their classes

We may expect students who do better in the course to do better on the FCI. If the majority of the students who took the FCI test were the ones who were struggling in the course and later failed the course or got poor grades, they might have lowered the learning gain as measured by the FCI. Hence, if a large fraction of students who took the FCI in earlier semesters after the reform (where the learning gain was higher) had received better grades than those taking the test in later semesters (where learning gain was lower) then it might explain the drop seen in the FCI gain in the later semesters.

We were able to match student FCI data to course grades (i.e. actual data of students

who got various grades in the course and took the FCI test) in 6 of the 8 semesters and not for 2 semesters, E and H, as their IDs were different. The Figure 5.2 shows the comparison of the various grades of the students who took the FCI to the overall students in the course for 6 semesters.

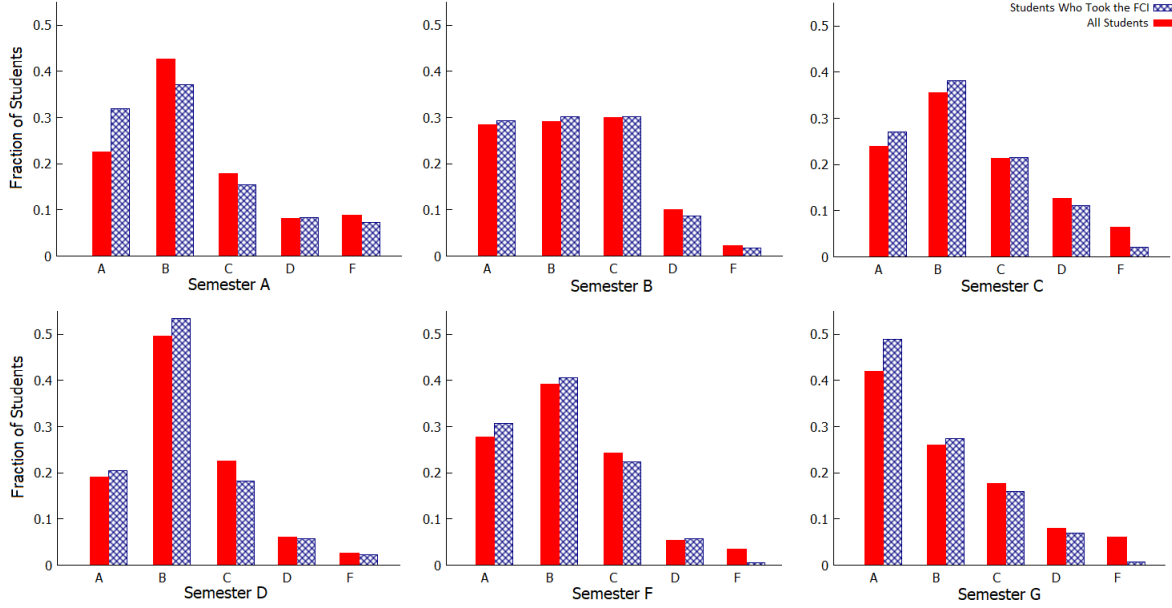


Figure 5.2: The bar graphs show the comparison of the fraction of students who got various grades (A,B,C,D, and F) in the course and took FCI to the fraction of the total students who got similar grades in the course for 6 semesters (A, B,C,D,F, and G). The data show that a higher percentage of students who got A grade in the course took the FCI test in the later semesters in comparison to the earlier semesters.

For each semester with matched data, the fraction of students who got grade A in the course and took FCI is slightly higher than the fraction of total students who got the same grade in the course. Similarly, the fractions of the students who got grades B or C and took FCI are nearly equal to the corresponding fractions of the total students who got similar grades in the course. Also, only small fractions of the students who got grades D and F took the FCI in each semester. Hence, the students who took the FCI seem to be the representative sample of the student population in the course as judged by final course grade. Moreover, a higher fraction of students who got grade A in the course took FCI in the later semesters compared to earlier semesters. Hence, the data suggest that the students

who took FCI represent their overall population in the classrooms and it is not that much higher fraction of students with poor grades took the FCI in the later semesters and thus dropped the learning gain.

Also, the FCI gain may depend indirectly upon the withdrawal rate of the students, assuming the struggling students are the ones who withdraw from the course. If fewer students are withdrawing from the course (which means at least some struggling students are continuing the course) in semesters with low FCI gain and more students are withdrawing (which means none or only a few struggling students are continuing the course) in semesters with higher FCI gain, then the students taking the FCI test in different semesters are academically at different levels. Hence, the drop in the fractional gain can be attributed to the variation of the withdrawal rate of the students. The largest withdrawal rate for all 8 semesters with FCI data is 3.6% of the total enrolled students. This number is too small to bring any appreciable changes in the FCI gain.

Clearly, the withdrawal rate of the students in any of these semesters is significantly low to produce any appreciable effect on the learning gain and the Figure 5.2 shows that slightly higher fractions of students who got grade A (and slightly smaller fractions of students who got grade C) in the course took the FCI test in later semesters than those in earlier semesters. Also, the bar graph of the students who took FCI is not that much different from the corresponding bar graph for the total students for all the grades and for all the semesters. This leads us to believe that the students taking the FCI in each semester are representative of the student population in the course and thus reject the hypothesis that high achievers took the FCI test in earlier semesters and low achievers took the FCI test in later semesters.

5.1.3 Hypothesis 3: Post-test scores are strongly influenced by when the FCI is taken

The FCI pre-test was given either at the first or the second meetings in the studio before the students were introduced to the Newtonian concepts in the lecture. Unfortunately, we do have data with two different timings of the FCI post-test. For the first four semesters,

it was given three quarters of the way through the semester (12th week of the semester) and for the last four semesters, it was given at the end of the semester (15th week of the semester). There is a difference of three weeks in between the timings of the two post-tests. There is a possibility that the students who took the FCI post-test three weeks earlier had better retention and therefore would have scored higher than the students who had to wait until the end of the semester to take the post-test. This difference in timing of the post-test could be the reason for the fall in the gain seen in later semesters.

At KSU, students take five mid-terms exams, at regular intervals, besides the final and none of these mid-term exams are cumulative. The topics contained in the FCI are the Newton's three laws, circular motion, kinematics and motion in one and higher dimensions²⁴. In every semester, all these topics are covered before the second test, i.e. within the first month of the beginning of the semester. Looking at the syllabus, at the three quarters of the semester, students learn oscillations and at the end of the semester, they learn thermodynamics. These topics are not covered in the FCI and are completely different from those covered. There is at least 50 days gap between the lecture of the circular motion, the last topics included in the FCI and the three quarters of the semester. Since the mid-term exams are not cumulative, we can presume that students do not revisit the topics covered in FCI after the second test until they prepare for the final.

Research done on physics problems related to Newton's third law and introductory electricity and magnetism shows that the percentage of correct responses to these problems increases either during the weeks of instruction or when students solve similar homework problems or when they take exam related to similar topics⁴⁵. After a couple of weeks, the percentage of correct responses levels off. In our case, at the three quarters of the semester or at the end of the semester when students took the FCI post-test, they were neither taught the related topics covered in FCI nor given any related homework problems. They also did not have to take an exam on related topics. Hence, it may not matter whether the FCI post-test was given three quarters of the way through the semester or at the end of the semester.

To provide further evidence that the timing of the post-test may not be the cause for the

drop in the fractional gain, we gave two post-tests to the students for a semester. The two post-tests were conducted at the 12th and the 15th weeks of the semesters. We analyzed the scores of 180 matched students who took all three FCI test, one pre and two posts, with ANOVA (with Bonferroni correction) and found that the scores for 12th and the 15th week were not statistically significantly different. Further details of week 12 vs week 15 are given in the Table 5.1.

Week	N	Post-Mean	Gain	Fractional Gain	SD	Cohen's d
12	180	0.580	0.123	0.235	0.114	1.06
15	180	0.543	0.078	0.159	0.131	0.74

Table 5.1: *Data Summary of Week 12 VS Week 15 FCI results*

Although there are two different timings of the post-tests in our data, it does not seem to have a significant impact on the learning gain, so may not be the reason for the fall in the learning gain as measured by FCI.

In summary, the extraneous effect that could have possibly lowered the FCI fractional gain in later semesters are dismissed, so is the drop in the FCI gain somehow related to the change in structure of the course, i.e. change in the curriculum and change in the method of instruction/instructors within the course?

5.2 Can we explain the changes in the FCI scores with changes in the course?

In the following section, we focus on the things that have changed over time. The components of the courses that have stayed relatively constant such as physical environment, laboratory materials, exams, and textbooks are safely ignored. We combine these changes according to the data that we have and then try to explain the drop in the fractional gain on the basis of these changes.

5.2.1 Combination of homework type and quiz/CGPS in the studio

Since fall 2008, Professor Circle has been using online homework instead of paper/written homework and CGPS instead of quiz in the studio. However, Professor Triangle still uses paper homework and quiz in the studio. For semesters with FCI data, online homework and CGPS were used in semesters E, F and H and for other semesters, paper homework and quiz were used. Whether its online or paper homework, students solve them twice a week.

The analysis of the gain using ANOVA with Bonferroni correction showed that the semester G is not statistically significantly different from the other semesters E, F and H. Hence, the combination of homework type and quiz/CGPS does not seem to have a significant effect on the learning gain. However, it is possible that one of them may be reinforcing the learning gain and the other may be diminishing the learning gain, but we cannot tell from the data that we have, so we refer to the available literature.

A research was done in an organic chemistry course with the undergraduate students to compare the content knowledge between students who completed online homework to those who completed paper homework. Students registered for one of the two sections of organic chemistry. The two sections were taught back to back in the same room by the same instructor with the same teaching method and content, including the same exam but with different homework methods. The two sections had students with identical average chemistry GPA. In one semester the online homework section had 26 students and the paper homework section had 36 students. In the subsequent semester, the online homework section had 20 students and the paper homework section had 26 students. Both sections received the homework with the same questions in the same layout and were given the same amount of time for completion. The content knowledge was measured by comparing students score on American Chemical Society Full-year organic exam which consists of 70 multiple choice questions. A one-tail t-test analysis showed that sections with online homework had statistically significant result compared to the section with paper homework implying that online homework enhanced student learning in organic chemistry⁴⁶.

Similarly, another research was done in a general chemistry course with two different student groups, one doing online homework ($N=45$) and the other doing paper homework ($N=45$). Weekly problem sets from the end of the textbook were given to both groups. Analysis using Mann-Whitney test showed that there was no statistically significant difference between the two groups suggesting that the online homework was equally as effective as paper homework for student learning as measured by a short but comprehensive exam that was developed by the authors for use as a pretest/post-test⁴⁷.

Another research to compare the effects of online homework with paper-based homework was done with physics students at a public university for 3 semesters ($N_{\text{paper}}=143$ for 3 semesters and $N_{\text{online}}=144$ for 3 semesters). Both groups were given identical problems as homework. The t-test analysis showed that there was no any statistically significant difference in exam scores and concept test as measured by CSEM between the two groups throughout the three semesters⁴⁸.

Similarly another research was done with physics students (110 students in each of the two sections of calculus-based introductory physics course and 60 students in each of the two sections of algebra-based introductory physics course) using two different homework methods (online and paper) but keeping all other parameters the same as far as possible (the same instructor teaching both sections back-to-back on the same day, essentially the same assignments were given to both sections). Student learning was measured by the exam and FMCE. The t-test analysis showed no statistically significant difference in either in the exam scores or in the FMCE gain⁴⁹.

All these published literature suggest that the use of online homework does not seem to have a significant impact on the learning gain, so the use of online homework may not be the reason for the decrease in the learning gain as measured by the FCI.

Similarly, we present the published literature on cooperative group problem solving that is related to our work. A research was done in an algebra-based introductory physics course to test if the solution to context-rich physics problems by students working in a collaborative group is better than the solution of the best student in the group. The students were given the seven tests besides the final exam, which was to be done independently. Each test was given

in two parts. The first part consisted of a context-rich problem to be solved in a cooperative group in 50 minutes and the second part consisted of a short qualitative question and two context-rich problems to be solved independently in 50 minutes the following day. Students turned in only one solution for the group problem and each member in the group received the same score. To determine whether one set of problem solution was better than another, the authors defined the better solution as the one that exhibited characteristics similar to the solution produced by an expert. Analysis of the scores of the group solution and the best-in-group individual solution with Wilcoxon Rank-Sum test showed that the group solutions were statistically significantly better than the best-in-group individual solution⁵⁰.

Similarly, another research was done with the algebra-based physics course to measure the effectiveness of in-class student-student interaction on the learning of physics. Two sections of the first semester of the course were taught by the same instructor keeping the syllabus, homework, exams, quizzes, and lab experiments identical for comparison. The lecture usually consisted of a short lecture and a set of worksheets. In one of the sections, the students had to solve the worksheet independently and after some time, the instructor solved the problem. Whereas, in the other section, the students were encouraged to discuss the problems in a group of two or three before calling on the instructor whose role was to go from one group to the other and help them accordingly. Each group was then asked to share the ideas and the approach to the final solution with the class and then the instructor presented the solution to the students if there was any confusion. Mechanics diagnostic pre/post-test, classroom examination, and student feedback were used for the assessment of the research. The t-test analysis of the scores of the Mechanics diagnostic test and classroom examination showed no statistically significant difference between the two groups of students. The results showed that the student-student interaction lead to an improvement of attitudes of the students towards the course, improved the academic environment of the class and made students feel better about the material they were learning. However, these qualitative improvements in the class did not lead to a significant change in the test scores on either the classroom test or the Halloun-Hestenes mechanics diagnostic test⁵¹.

These published literature do not suggest that CGPS has a significant negative impact

on the learning gain, so the introduction of CGPS may not seem to lower the learning gain as measured by the FCI. Our data and related published literature suggest that the introduction of online homework and the CGPS in the studio do not seem likely to be the reasons for the decline in the learning gain as measured by the FCI.

5.2.2 Changes in the studio instructors

As mentioned in the chapter 4, the studio instructors for the earlier semesters were either the faculty or the post-docs and the secondary instructors were either the graduate student or the senior undergrad. However, in the later semesters graduate students taught as the primary instructors and LAs were also introduced to assist as the secondary instructors. Are the graduate students and the LAs responsible for the drop in the learning gain?

1. Graduate Student as Studio Primary

Although graduate students were working as the primary instructors in the studio much before, for the semester with FCI data, they were teaching as studio primary since semester E. Is the introduction of the graduate students as studio primary responsible for the fall in the learning gain? To answer this question, we ran ANOVA with Bonferroni on Gain across all instructor types and found that the graduate students as primary instructors are not statistically significantly different from either the faculty or the instructors hired to teach as the primary instructors in the studio. We also plotted the gain, Cohens d, and fractional gain of the various studio instructors within a semester for different semesters since semester E and are shown by the Figure 5.4, Figure 5.3, and Figure 5.5.

Figure 5.3 and Figure 5.4 show that the values of Cohens d and the gain are comparable across instructor types and the sections led by graduate students seem to do better than some other groups. Also, statistical test reveals that the graduate students as primary instructors are comparable to the faculty teaching as the primary instructors in the studio. Hence, the introduction of graduate students as primary instructors is unlikely to have caused the drop in the learning gain.

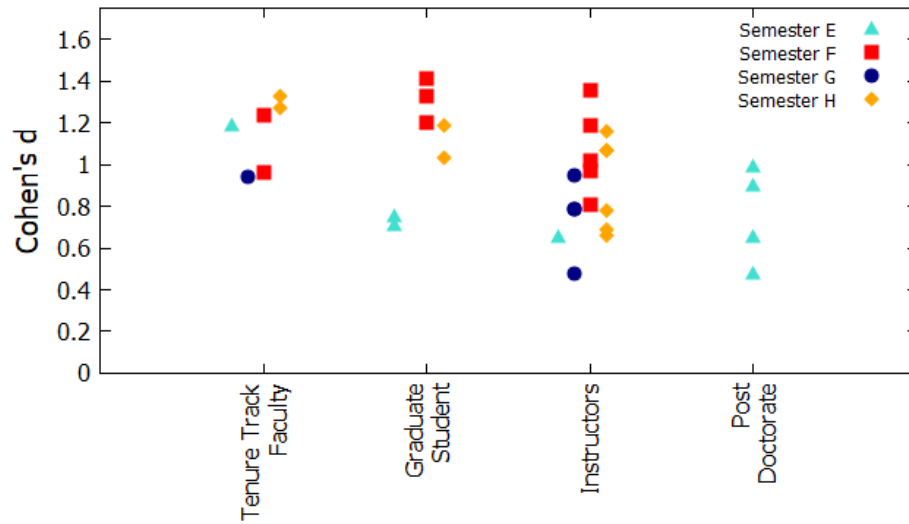


Figure 5.3: Variation of the Cohen's d of the various primary instructors for different studios in semesters E, F, G and H. Graduate students were working as primary instructors in semesters E, F, and H.

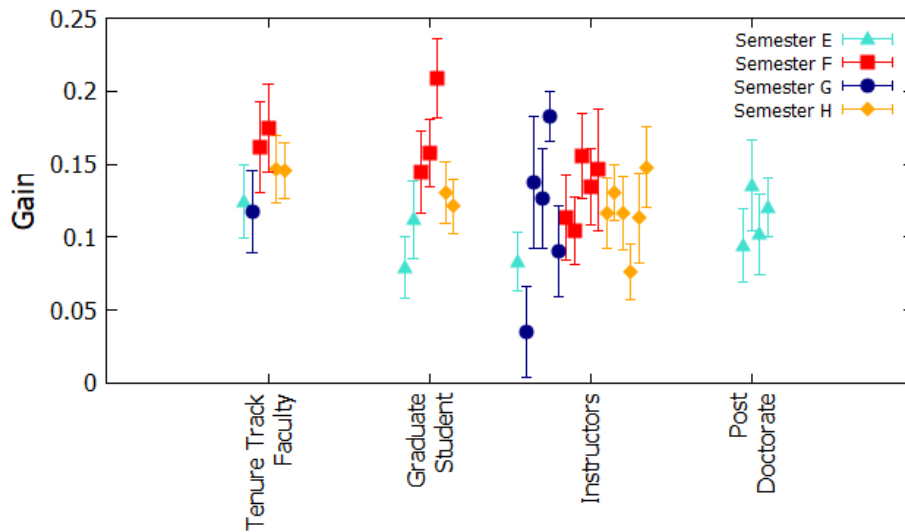


Figure 5.4: Variation of the gain of the various primary instructors for different studios in semesters E, F, G, and H. The error bars are the standard error of the mean. The variation of the gain of the studio sections with graduate students as primary instructors is similar to that of other groups and the gains not the lowest.

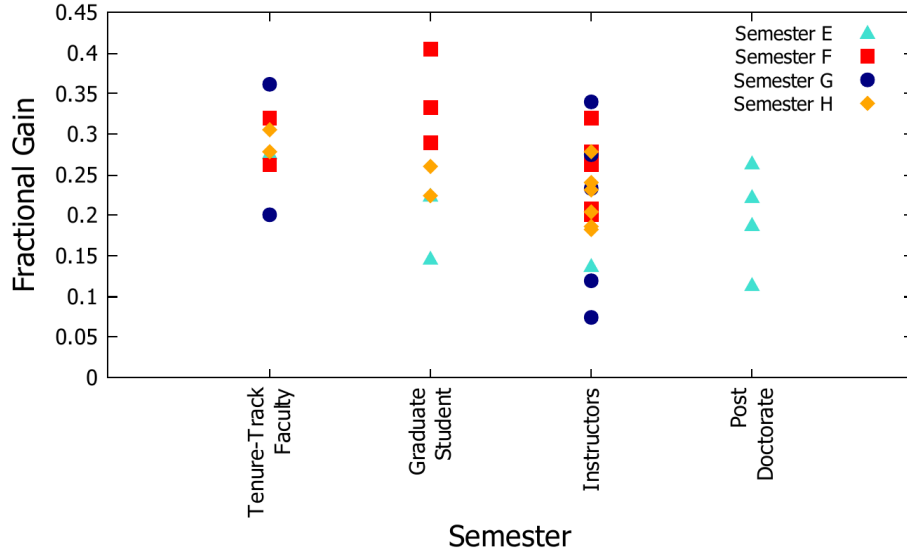


Figure 5.5: Variation of the fractional gain of the various primary instructors for different studios in semesters E, F, G, and H.

2. Learning Assistant as Secondary Instructors

Previous research from 17 institutions shows that LA supported programs are advantageous in enhancing student understanding of the core ideas of physics⁵². Similarly, other researches have demonstrated that the students have shown significantly increased learning gains in LA-facilitated courses^{35;53;54}. Since spring 2017, the physics department at KSU has introduced LAs in the studio. For semesters with FCI data, semesters F, G and H involve learning assistants in the studio. The ANOVA with Tukey test on the gain, using teaching assistant as the comparison parameter gives a p-value of 0.27. Hence, TA or LA working as the secondary instructors does not appear to be significant, the post-mean for the two different assistants (graduate teaching assistant and LAs) were found to be not statistically significant (For graduate teaching assistant, the post mean was 0.600 and that for the LA was 0.607).

Hence, the introduction of graduate students as the primary instructors and the LAs as the secondary instructors in the studio do not seem likely reasons for the drop in the learning gain as measured by the FCI.

5.2.3 Use of clickers in the lecture

Although clickers have been used in the lecture hall since Fall 2008, for semesters with FCI data, they are being used since semester E. The use of the clickers is expected to enhance active learning. While we have no direct way of checking to see if the introduction of clicker questions had a significant effect on the learning gain, we consider this unlikely given the existing literature on clicker questions. Here are some of the published literature on the use of clickers.

A study was done with 508 undergraduate students enrolled in an introductory management course dividing them into 2 groups, clickers and non-clickers class sections with the approximately the same number of students. The two groups had the same instructors, textbook, topic coverage, and homework assignments. The t-test analysis showed that the use of clickers had a statistically significant impact on open-ended exam questions but no measurable impact on multiple choice exam questions⁵⁵.

Another research was done with four sections (of about 45 to 55 students in each section) of microeconomics taught by two different instructors. Each instructor has one section using clickers and the other using no clickers. The exam scores of the students were compared between those who used the clickers to those who did not. Analysis (using t-test and educational production function developed by Hanushek⁵⁶ showed that after controlling for other factors, students in the clicker sections did no better or worse than those in the non-clicker sections⁵⁷.

Similarly, another research was done with multiple introductory courses (psychology, education, speech, and accounting) in a university to determine the effectiveness of the clickers. Students in ten sections of the introductory courses (four sections of psychology, two sections of accounting, two sections of education, and two sections of speech) agreed to participate in the study. Five faculty were the instructors for these selected courses, each teaching two sections, one with the clickers and the other without the clickers. The assignments, materials covered, exams, and readings were similar as possible in both the sections for each instructor. Statistical analysis using the chi-squared test gave no statistically significant difference in

the course grade for these two different sections⁵⁸.

Yet another study was done with the biochemistry course (BIOCH 200) in Fall 2007 in a university to investigate the effect of using clickers in the classroom on the examination scores and to evaluate the perception of the students regarding clickers and their effectiveness in the classrooms. The three different sections (A1, A2, and A3) of the course were offered, one (A1) that used clickers in the lecture and the other two (A2 and A3) did not use the clickers. Students in the clickers section (A1 of Fall 2007) were compared with two other sections (A2 and A3 of Fall 2007) and all other sections of the same course before the implementation of the clickers (4 sections of Fall 2005, 3 sections of Winter 2006, 3 sections of Fall 2006, and 3 sections of Winter 2007). The range of the mean scores for 13 sections of the course completed before the study was 68.76-75.80%.

“All students regardless of the sections had access to identical learning materials, including recommended readings from the text, instructors’ slide files, highly specified learning objectives and practice questions. All instructors worked from the same extensive and detailed set of instructors’ notes. In each term, the students took the same exam at the same time. The questions for the exam in each term were drawn from the large set of database, and all questions in the database had difficulty and reliability index associated with them”.

The results showed that the students in the clickers section had a slightly higher mean composite examination score. However, the range of mean composite scores for all the three sections of Fall 2007 were within the range of mean composite score that was observed previously for similar examinations, so the researchers concluded “overall, there was no apparent difference in the mean composite examination score for students taught with clickers compared with those taught by traditional lectures. Also, the researchers observed that greater than the usual number of students in the clickers section had got the grade of A or A+ (composite score 87%). The result of the survey showed “the majority of the students agreed that the clicker questions used in class helped them to improve their learning (86%), to build a solid understanding of core concepts (85%), to focus and pay more attention in lectures (80%), and to engage directly with the content being presented (87%). The majority agreed also that they felt they had an advantage in the examinations over the

other two sections of the course because of the clicker questions (83%), and that the use of clickers should be continued in this course in the future (84%)⁵⁹.

Also, another research was done in a calculus-based introductory physics course to find the effect of using question sequences with voting machines on student learning and their perception of its effect on their learning. Two different sections of a similar population of students with the same course content, homework, recitation, and the labs were compared. One section used the voting machine and the other section did not. Students in the class with voting section were encouraged to discuss the questions with each other before voting. The analysis of the data from CSEM pre/post-test and common exam questions using t-test showed that students with voting machine section achieved a small but significant gain in conceptual learning. The attitude survey revealed that the students were strongly positive regarding the use of the voting machine and believed that it helped them learn better⁶⁰.

Similarly, another research to measure the effectiveness of the clickers and the perception of the students was done with the architecture students in Fall 2008 and their performance was compared to a control group of students in Fall 2007 and Fall 2006. The instructor used the same textbook and the lecture material for all these years. The t-test analysis showed that the students had significantly better grades in Fall 2008 compared to those in Fall 2007 and Fall 2006. The students reported that clickers were fun, they increased class participation and involvement, were effective to their learning experience and help keep them feeling mentally engaged, academically responsible and accountable⁶¹.

All these published literature suggest that the use of clickers in the classroom may not seem likely for the decrease in the learning gain as measured by the FCI in the later semesters.

Chapter 6

Discussion, Conclusion, and Future Work

Educational researchers in principle agree that students have difficulties in learning science because of the passive role the students have in the traditional lectures⁶² and considering the volume of the research work in the active learning, it is well documented and accepted that active learning substantially increases the learning gain¹⁸⁻²⁰. However, relatively recent research done in biology with randomly selected multiple institutions across 22 states revealed that there was no correlation between student learning and the use of active learning instructions⁶³. At KSU, previous work^{3;30} confirms that the reform of the calculus-based introductory physics course provided better learning gain, as measured by the FCI, immediately after the reform. However, our work reveals that the learning gain in the later semesters decreased and is comparable to that of the traditional methods of instructions although active learning methods of instruction are still being used. In fact, in the later semesters, the course involves more active learning such as the implementation of CGPS and the use of clickers in the lectures, and there has not been a substantial change in the curriculum. Therefore, we do not expect the drop in the learning gain in the later semesters. Our work reveals that incorporating active learning into the curriculum simply does not guarantee an enhancement in student understanding as measured by the concept inventories and sustain-

ing a reformed course is challenging. We cannot take for granted that a reformed course that once provided better learning gain will provide at least the same level of learning gain for enduring future.

The drop in the learning gain as measured by the FCI does not mean we need to revert to the traditional system of lecture/lab/recitation as there are likely other benefits (such as the improvement in the students' quantitative problem-solving skills, improvement in critical thinking ability, development of collaborative skills, improvement in retention, improvement in communication skills, mathematical skills, and positive shifts in students' attitude towards physics) of the active learning system. However, we were not able to compare any one of them over time as they were not the part of the original research effort.

6.1 Limitation

The major source of our information is the syllabi and we have only anecdotal information about what actually happened in the classrooms. Our research is mostly focused on the FCI and DFW data and we have no information about how the instructors in different semesters taught the topics covered in FCI. We also do not know if any lead instructors directly taught the questions of the FCI or asked the same questions on the exams.

The area in which there seems to be the most change over the years, instruction, is also the area where we know the least about the course. While we know a lot about who the instructors are and what the lead instructor intends (by looking at the syllabus) for the method of instruction to be like each primary studio instructor may teach their studio section slightly differently from the next and it is in the studio where most of the active learning is involved and two third of the time is spent. Maybe some of them spend a significant amount of time reviewing material and have less time for the experiment or possibly they choose not to do a review of the homework during the class. Considering the volume of studio primaries, secondaries and studio sections that exist, it is infeasible to keep track of the differences between primaries' methods of instruction and variations in the performance of TAs as there has been no explicit TA training, that could provide consistency in performance

across all studios, since the reform.

Most importantly, we want to reiterate that our data collection took place near the end of this project, thus restricting the kind of information that we could possibly obtain. These variations in the instructional decisions among the TAs, studio primaries and the lead instructors themselves might have resulted in the variation in the learning gain. A case study that examined the variations in physics faculty practices in the secondary implementation of peer instruction in introductory physics courses revealed that the instructors actual practice in the classroom differ strikingly resulting in large discrepancies in students' opportunities to engage in formulating and asking questions, evaluating the correctness and completeness of problem solutions, interacting with physicists and communicating scientific ideas⁶⁴.

6.2 Suggested Explanation

6.2.1 Newness Effect

One possible explanation for the drop in the learning gain in the later semesters is the idea that the faculty were more engaged in the pedagogy and were more thoughtful in teaching the course immediately after the reform. Moreover, the lead instructors for semesters immediately after the reform were the ones who were directly involved in the planning and design of the reform, so they might have a more rich and nuanced understanding of what they were doing and how the curriculum was set up. The excitement the reform brought, the involvement the reform required for its successful implementations and the dire consequences, if any, of its failure, might have provided them the extra motivation and dedication for its successful implementation. Their increased interest in the course could have also increased their engagement in the material and overall efforts when teaching the course. Moreover, the initial interviews with the instructors teaching the course, including the studios, revealed that they had positive reactions towards the change and felt that the new format was a better method of teaching as compared to the previous method and that the students were learning better in the new format.

Also, from the students' point of view, they had ideas of how the course was taught before the reform from their friends and colleagues who had taken the course before³. Thus, the students knew that it was a new program and thus might have behaved differently which might have resulted in better learning gains. The initial interviews with the students about their perception of the new studio and the course overall had revealed that they were, in most parts, positive about the changes done in the course. They liked the integration of the lab with the discussion of the homework problems, the opportunity of getting individual attention and care in the classroom setting and the interactions they had with their peers. They felt that there were connections among the lecture, homework and the laboratory experiments. These positive attitudes towards the course and the satisfaction in taking the course might have enhanced their motivation which might have contributed to better understanding and higher learning gains in earlier semesters.

Now, we are almost two decades after the reform of the course and it is possible that this newness effect has worn off as there is no more excitement among the instructors and the students regarding the course structure. Aside from minor adaptations such as the use of clickers during the lecture or the introduction of LAs in the studio, the process of teaching the course using the interactive-engagement method is now just "the way things are". That original excitement has dissipated.

It is also valuable to note that since none of the original lead instructors from the semesters directly following the reform are still acting as lead instructors, they might have passed the information along to the new lead instructors. This could be looked at like a game of "telephone". One instructor must explain the process to the next instructor, but it is inevitable that what the next instructor has heard is slightly different than how the original instructor meant it. When this information has been passed through the hands of several different people for almost two decades, there are bound to be differences that evolve simply from the interpretation of the information by each individual. Also, as the information is passed on, details of certain things related to the studio or the lecture might get lost and this might have drifted the studio/lecture from the original intentions possibly resulting to poor learning gains.

6.2.2 Group Formation in the Studio

While we have been unable to confirm this with documentation, we have anecdotal evidence that there have been changes in the way the students' groups are formed in the studio. In earlier semesters, when there was higher FCI fractional gain, the groups were very carefully constructed after each exam. The group consisted of the students of mixed abilities (high, medium and low performers on the exam) but in the later semesters, the groups were randomly formed. For example, the students are asked to pick a numbered card and were told to move to the corresponding table.

In a group of mixed abilities, the weaker ones can get assistance from the stronger ones who can also learn by teaching others⁵⁰. Whereas in a randomly formed group, some of the groups might not be well functioning as there are possibilities that some groups might have consisted of only low achievers and asking them to work together might not have resulted in enough learning gain as was anticipated in a team effort. For instance, the well-constructed groups in physics II have brought less variance in performance than the randomly generated groups in Physics I⁶⁵. Hence, the change in the way groups are formed in the studio may be the reason for the drop in the learning gain in the later semesters.

It could be that if we had more insight into these aspects of the course, then we would be able to find the answer to our question of what has caused the drop in FCI gain. At Harvard, introducing peer instruction into calculus-based introductory physics course increased the FCI fractional gain monotonically from 1990 to 1997²⁸. However, they had made some notable changes in the curriculum since 1993 and most importantly peer instruction was developed at Harvard, so the primary implementation might have a rich and nuanced understanding of the reform better than the secondary implementations like ours. This suggests that one of the ways of maintaining higher learning gain is to consistently improve the curriculum according to the instructional design set up.

As suggested by Hake and discussed in chapter 1 of the thesis, our data show variation in the learning gain that is not uncommon even within the active learning courses, either in the primary implementation or in the secondary implementation of the reform¹⁹. The issue

that we have is that the learning gain was higher initially after the reform (normalized gain was 0.41, 0.41 and 0.39) but in the later semesters, it varied from 0.19 to 0.27, which is low for the active engagement course and we don't know exactly what caused the drop although we have eliminated several possible explanations for the drop.

Ideally, we want the reform to provide better learning gain for many years and want the reformed system to be enduring and sustainable for years to come as the reform of any course requires a lot of resources in terms of person, money, time, and efforts. Being unable to find a concrete explanation for the drop in FCI fractional gain in the later semesters, we consider other possible explanation, for future studies.

6.3 Recommendation for Future Research Work

For future study, one may look at the amount of time faculty spend thinking about teaching and how that changes over time as a measure of investment in the course; study the alignment of the curriculum, instruction, and assessment and how it changes over time; how do students' perception, motivation and satisfaction in taking the course change over time. One may also study how a reformed course changes when it is passed off to a new faculty who is not involved in the original development.

One can also study how the group formation in the studio affects the learning gain. One may form two type of groups; one that consists of students with mixed abilities, as recommended by the original developers of CGPS, and the other randomly formed groups and study both types of groups. If these groups perform significantly differently (with the carefully chosen groups outperforming the randomly selected groups) then the random formation of the groups of the students in the studio might have been hindering the students from learning basic Newtonian concepts as measured by the FCI.

In the future, one may not use not only the concept inventories but also the quantitative problem- solving standard tests such as Minnesota Assessment of Problem Solving (MAPS) or Assessment of Textbook Problem Solving Ability (ATPSA) to assess how students' over-all learning is changing with time. Also, as the attitude of the students plays a big role in

learning, one may ask them to take the Maryland Physics Expectation (MPEX) Surveys or the Colorado Learning Attitudes about Science Survey (CLASS) and analyze their scores together with the FCI scores and the scores from the quantitative problem solving assessments to see if the change in the learning gain can be related to the changes in the attitudes of the students if there is any. Assessment using concept inventories, attitude test, and quantitative problems may provide detail and enough evidence of how successful the reform has been and how helpful it has been in helping the students learn physics.

In the future, we may take videos of the lectures and the studios and analyze them to see how the lectures of different lead instructors differ, how much time does each lead instructor spend in the materials covered in the FCI, how effectively do different lead instructors use the clicker questions in the lectures, how seriously do students answer the clicker questions, how do TAs and the primary instructors in different studio teach, how much do students in different studio sections involve themselves in the active learning, how much do each students in a cooperative group contribute in learning or does someone just sit there and wait for the others to provide them the answers, how is the distribution of time for the quiz/CGPS, lab experiment and the review of the homework problems in each studio, how much do the students actually get the opportunities to engage in active learning and how much does the teaching strategies differ from that of the original intentions. All these detail video analyses can provide answers to the changes in the learning gain if there is any in the future.

6.4 Conclusion

In this thesis, we studied what happened after the reform, of the first semester of the calculus-based introductory physics course, that took place almost 20 years ago. We looked at how the physical environment, curriculum, the method of instruction, DFW rate, and learning gain as measured by the FCI have changed since the reform. Previous research work^{3;30} had shown that the reform was successful in improving the learning of the students. The system that was put in place by the originators of the reform seems to be self-sustaining because the overall structure of the reformed course has remained largely intact.

For the physical environment, we found that it has remained almost the same with the exception that more rooms have been used in the later semesters to accommodate the influx of students.

For the curricular materials, we found that the course content has remained the same. The students have been doing almost the same experiments since the reform of the course but there have been some modifications such as the introduction of the cooperative group problem solving and the use of online homework by some lead instructors teaching this course. Also, only three different textbooks have been used so far and one of them was being used before the reform of the course.

Regarding the instructors, we found that there have been some changes in instructors such as the introduction of the graduate students as primary instructors and LAs in the studio. For the method of instruction, we found that the clicker questions have been used in the lecture halls since its introduction about a decade since the reform took place.

We also found that there is a large variation in the DFW rate from one semester to another but the trendline is constant and lower compared to before the reform. We used FCI to assess the learning gain of the student and the success of the reform. The data show that the learning gain was higher immediately after the reform, but it dropped in the later semesters for which we have the FCI scores. We showed that our data is reliable and the fall in learning gain is not caused by

1. better preparation of the students coming into the course in earlier semesters.
2. changes in the performance levels of students who took the test.
3. the difference in timing of the post-test that we have in our data.

We examined several possible explanations for the fall in the learning gain as measured by FCI. However, the changes made in the course structure and in the method of instruction, such as

1. combination of homework type and quiz/CGPS in the studio do not explain the drop in the learning gain.

2. introduction of graduate students as primary instructors and LAs as secondary instructors do not explain the drop in the learning gain.
3. use of clickers in the lecture does not explain the drop in the learning gain.

In fact, in later semesters, the course involves more active learning such as the use of clickers and CGPS (except for semester G) which are expected to enhance student learning.

The focus of the research has been to explore the reasons for the fall in the learning gains in the later semesters. Knowing why it dropped would give us information about how to sustain future reforms. This would help ensure that future investments in reform efforts remained fruitful over longer periods of time. We were able to eliminate several possible explanations for the drop in the learning gain.

Although there has been a drop in the learning gain, the reform has been successful in maintaining a lower DFW rate. The drop in the learning gain as measured by the FCI does not mean we need to revert to the traditional system of lecture/lab/recitation as there are likely other benefits of the active learning system that are not being measured. Hence, the reform may have benefited the students because of the active learning involved and the organization of the course.

Further research in the reformed courses needs to be done to come up with a more generalized idea of how a successful reformed course would look after many years and how can we sustain reform for longer periods of time so that we continue getting the benefits of the massive investments made in the reform.

Bibliography

- [1] Robert J Beichner, Jeffery M Saul, David S Abbott, Jeanne J Morse, Duane Deardorff, Rhett J Allain, Scott W Bonham, Melissa H Dancy, and John S Risley. The student-centered activities for large enrollment undergraduate programs (scale-up) project. *Research-based reform of university physics*, 1(1):2–39, 2007.
- [2] Eric Brewe, Remy Dou, and Robert Shand. Costs of success: Financial implications of implementation of active learning in introductory physics courses for students and administrators. *Phys. Rev. Phys. Educ. Res.*, 14:010109, Feb 2018. doi: 10.1103/PhysRevPhysEducRes.14.010109. URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.14.010109>.
- [3] Alice Churukian. *Interactive Engagement in an Introductory University Physics Course: Learning Gains and Perceptions*. PhD thesis, Kansas State University, 2002.
- [4] Melissa Dancy and Charles Henderson. Barriers and promises in stem reform. In *National Academies of Science Promising Practices Workshop*, 2008.
- [5] Charles Henderson and Melissa H. Dancy. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Phys. Rev. ST Phys. Educ. Res.*, 3:020102, Sep 2007. doi: 10.1103/PhysRevSTPER.3.020102. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.3.020102>.
- [6] Charles Henderson and Melissa H Dancy. Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics*, 76(1):79–91, 2008.
- [7] Kathleen Foote, Alexis Knaub, Charles Henderson, Melissa Dancy, and Robert J. Beichner. Enabling and challenging factors in institutional reform: The case of scale-up.

- Phys. Rev. Phys. Educ. Res.*, 12:010103, Feb 2016. doi: 10.1103/PhysRevPhysEducRes.12.010103. URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.010103>.
- [8] Manjula D. Sharma, Ian D. Johnston, Helen Johnston, Kevin Varvell, Gordon Robertson, Andrew Hopkins, Chris Stewart, Ian Cooper, and Ronald Thornton. Use of interactive lecture demonstrations: A ten year study. *Phys. Rev. ST Phys. Educ. Res.*, 6:020119, Oct 2010. doi: 10.1103/PhysRevSTPER.6.020119. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.6.020119>.
- [9] Charles Henderson, Melissa Dancy, and Magdalena Niewiadomska-Bugaj. Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics-Physics Education Research*, 8(2):020104, 2012.
- [10] John W Belcher. Improving student understanding with teal. *Faculty Newsletters*, 16(8), 2003.
- [11] Elaine Seymour. Tracking the processes of change in us undergraduate education in science, mathematics, engineering, and technology. *Science Education*, 86(1):79–105, 2002.
- [12] Ibrahim Abou Halloun and David Hestenes. The initial knowledge state of college physics students. *American Journal of Physics*, 53(11):1043–1055, 1985. doi: 10.1119/1.14030. URL <https://doi.org/10.1119/1.14030>.
- [13] Sheila Tobias. *They're not Dumb, They're different: Stalking the second tier*. Research Corporation, 1994.
- [14] David P Maloney. Research on problem solving: Physics. *Handbook of research on science teaching and learning*, pages 327–354, 1994.
- [15] J Handelsman, D Ebert-May, R Beichner, P Bruns, A Chang, R DeHaan, et al. Scientific teaching. science. *Journal of science education*, 3(4):52–59, 2004.

- [16] Lillian Christie McDermott. Oersted medal lecture 2001: physics education research the key to student learning. *American Journal of Physics*, 69(11):1127–1137, 2001. doi: 10.1119/1.1389280. URL <https://doi.org/10.1119/1.1389280>.
- [17] Eric Mazur. *Peer Instruction: A User's Manual*. Pearson, Upper Saddle River, NJ, 1 edition edition, August 1996. ISBN 978-0-13-565441-5.
- [18] S Freeman, S L Eddy, M McDonough, M K Smith, N Okoroafor, H Jordt, and M.P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, May 2014.
- [19] Richard R Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1):64–74, January 1998.
- [20] Joshua Von Korff, Benjamin Archibeque, K. Alison Gomez, Tyrel Heckendorf, Sarah B. McKagan, Eleanor C. Sayre, Edward W. Schenk, Chase Shepherd, and Lane Sorell. Secondary analysis of teaching methods in introductory physics: A 50k-student study. *American Journal of Physics*, 84(12):969–974, November 2016. ISSN 0002-9505. doi: 10.1119/1.4964354. URL <http://aapt.scitation.org/doi/full/10.1119/1.4964354>.
- [21] National Research Council. *Adapting to a Changing World Challenges and Opportunities in Undergraduate Physics Education*. National Academies Press, 2013.
- [22] Jack M. Wilson. The couple physics studio. *The Physics Teacher*, 32(9):518–523, 1994. doi: 10.1119/1.2344100. URL <https://doi.org/10.1119/1.2344100>.
- [23] Charles Henderson and Melissa H. Dancy. Impact of physics education research on the teaching of introductory quantitative physics in the united states. *Phys. Rev. ST Phys. Educ. Res.*, 5:020107, Dec 2009. doi: 10.1103/PhysRevSTPER.5.020107. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.5.020107>.
- [24] David Hestenes, Malcolm Wells, and Gregg Swackhamer. Force concept inventory. *The Physics Teacher*, 30(3):141–158, March 1992.

- [25] Ronald K Thornton. Rk thornton and dr sokoloff, assessing student learning of newtons laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *am. j. phys.* 66, 338 (1998). *Am. J. Phys.*, 66:338, 1998.
- [26] R Chabay and B Sherwood. Brief electricity and magnetism assessment. *Physical Review Special Topics-Physics Education Research*, 2(1):7–13, 2006.
- [27] 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. doi: 10.1119/1.1371296. URL <https://doi.org/10.1119/1.1371296>.
- [28] Catherine H Crouch and Eric Mazur. Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9):970–977, 2001.
- [29] Wendell Potter, David Webb, Cassandra Paul, Emily West, Mark Bowen, Brenda Weiss, Lawrence Coleman, and Charles De Leone. Sixteen years of collaborative learning through active sense-making in physics (clasp) at uc davis. *American Journal of Physics*, 82(2):153–163, 2014. doi: 10.1119/1.4857435. URL <https://doi.org/10.1119/1.4857435>.
- [30] C. M. Sorensen, A. D. Churukian, S. Maleki, and D. A. Zollman. The new studio format for instruction of introductory physics. *American Journal of Physics*, 74(12):1077–1082, 2006. doi: 10.1119/1.2358999. URL <https://doi.org/10.1119/1.2358999>.
- [31] Matthew A. Kohlmyer, Marcos D. Caballero, Richard Catrambone, Ruth W. Chabay, Lin Ding, Mark P. Haugan, M. Jackson Marr, Bruce A. Sherwood, and Michael F. Schatz. Tale of two curricula: The performance of 2000 students in introductory electromagnetism. *Phys. Rev. ST Phys. Educ. Res.*, 5:020105, Oct 2009. doi: 10.1103/PhysRevSTPER.5.020105. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.5.020105>.

- [32] Law Priscilla. *Workshop Physics Activity Guide, The Core Volume with Module 1: Mechanics I: Kinematics and Newtonian Dynamics (Units 1-7)*. Wiley, 2004.
- [33] JM Saul and EF Redish. Final evaluation report for fipse grant# p116p50026: Evaluation of the workshop physics dissemination project. *University of Maryland*, 1997.
- [34] Shaffer McDermott. *Tutorials in Introductory Physics*. Pearson College Div, 2001.
- [35] Steven J. Pollock and Noah D. Finkelstein. Sustaining educational reforms in introductory physics. *Phys. Rev. ST Phys. Educ. Res.*, 4:010110, Jun 2008. doi: 10.1103/PhysRevSTPER.4.010110. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.4.010110>.
- [36] Barbara K Curry. *Instituting Enduring Innovations. Achieving Continuity of Change in Higher Education*. ERIC, 1992.
- [37] Adrianna Kezar. Synthesis of scholarship on change in higher education. 2009.
- [38] Carol L Colbeck. Assessing institutionalization of curricular and pedagogical reforms. *Research in Higher Education*, 43(4):397–421, 2002.
- [39] Charles Henderson, Andrea Beach, and Noah Finkelstein. Facilitating change in undergraduate stem instructional practices: An analytic review of the literature. *Journal of research in science teaching*, 48(8):952–984, 2011.
- [40] Maura Borrego and Charles Henderson. Increasing the use of evidence-based teaching in stem higher education: A comparison of eight change strategies. *Journal of Engineering Education*, 103(2):220–252, 2014.
- [41] Valerie Otero, Steven Pollock, and Noah Finkelstein. A physics departments role in preparing physics teachers: The colorado learning assistant model. *American Journal of Physics*, 78(11):1218–1224, 2010. doi: 10.1119/1.3471291. URL <https://doi.org/10.1119/1.3471291>.

- [42] Patricia Heller, Ronald Keith, and Scott Anderson. Teaching problem solving through cooperative grouping. part 1: Group versus individual problem solving. *American Journal of Physics*, 60(7):627–636, 1992. doi: 10.1119/1.17117. URL <https://doi.org/10.1119/1.17117>.
- [43] 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. doi: 10.1119/1.17118. URL <https://doi.org/10.1119/1.17118>.
- [44] Vincent P. Coletta, Jeffrey A. Phillips, and Jeffrey J. Steinert. Interpreting force concept inventory scores: Normalized gain and sat scores. *Phys. Rev. ST Phys. Educ. Res.*, 3:010106, May 2007. doi: 10.1103/PhysRevSTPER.3.010106. URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.3.010106>.
- [45] Andrew F. Heckler and Eleanor C. Sayre. What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course. *American Journal of Physics*, 78(7):768–777, June 2010. ISSN 0002-9505. doi: 10.1119/1.3384261. URL <https://aapt.scitation.org/doi/10.1119/1.3384261>.
- [46] Kinza Malik, Nylvia Martinez, Juan Romero, Skyler Schubel, and Philip A Janowicz. Mixed-methods study of online and written organic chemistry homework. *Journal of Chemical Education*, 91(11):1804–1809, 2014.
- [47] Herb Fynewever. A comparison of the effectiveness of web-based and paper-based homework for general chemistry. *The Chemical Educator*, 13(4):264–269, 2008.
- [48] Tolga Gok. Comparison of student performance using web-and paper-based homework in large enrollment introductory physics courses. *International Journal of Physical Sciences*, 6(15):3778–3784, 2011.
- [49] Scott Bonham, Robert Beichner, and Duane Deardorff. Online homework: Does it make a difference? *The Physics Teacher*, 39(5):293–296, 2001.

- [50] 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.
- [51] Mohammad Samiullah. Effect of in-class student–student interaction on the learning of physics in a college physics course. *American Journal of Physics*, 63(10):944–950, 1995.
- [52] Jada-Simone S White, Ben Van Dusen, and Edward A Roualdes. The impacts of learning assistants on student learning of physics. *arXiv preprint arXiv:1607.07469*, 2016.
- [53] Renee Michelle Goertzen, Eric Brewster, Laird H Kramer, Leanne Wells, and David Jones. Moving toward change: Institutionalizing reform through implementation of the learning assistant model and open source tutorials. *Physical Review Special Topics-Physics Education Research*, 7(2):020105, 2011.
- [54] Paul M Miller, Jeffrey S Carver, Aniketa Shinde, Betsy Ratcliff, and Ashley N Murphy. Initial replication results of learning assistants in university physics. In *AIP Conference Proceedings*, volume 1513, pages 30–33. AIP, 2013.
- [55] Maureen Lojo. A comparison of clicker effectiveness for multiple choice and quantitative questions. *California Journal of Operations Management*, 7(1):130–135, 2009.
- [56] Eric A Hanushek. Conceptual and empirical issues in the estimation of educational production functions. *Journal of human Resources*, pages 351–388, 1979.
- [57] Marianne Johnson, Denise Robson, et al. Clickers, student engagement and performance in an introductory economics course: A cautionary tale. *CHEER Virtual Edition*, 20: 4–12, 2008.
- [58] Robin K Morgan. Exploring the pedagogical effectiveness of clickers. *Insight: A journal of scholarly teaching*, 3:31–36, 2008.

- [59] Stephen Addison, Adrienne Wright, and Rachel Milner. Using clickers to improve student engagement and performance in an introductory biochemistry class. *Biochemistry and Molecular Biology Education*, 37(2):84–91, 2009.
- [60] Neville W Reay, Pengfei Li, and Lei Bao. Testing a new voting machine question methodology. *American Journal of Physics*, 76(2):171–178, 2008.
- [61] Leonard Bachman and Christine Bachman. A study of classroom response system clickers: Increasing student engagement and performance in a large undergraduate lecture class on architectural research. *Journal of Interactive Learning Research*, 22(1):5–21, 2011.
- [62] CC Bonwell and J Eison. Active learning: Creating excitement in the classroom (asheric higher education report no. 1) washington. *DC: George Washington University*, 1991.
- [63] Tessa M Andrews, Michael J Leonard, Clinton A Colgrove, and Steven T Kalinowski. Active learning not associated with student learning in a random sample of college biology courses. *CBELife Sciences Education*, 10(4):394–405, 2011.
- [64] Chandra Turpen and Noah D Finkelstein. Not all interactive engagement is the same: variations in physics professors implementation of peer instruction. *Physical Review Special Topics-Physics Education Research*, 5(2):020101, 2009.
- [65] Patrick B Kohl and H Vincent Kuo. Chronicling a successful secondary implementation of studio physics. *American Journal of Physics*, 80(9):832–839, 2012.