Teaching-as-Research Projects at CU Boulder

A Volume of Projects Done by the Teaching Institute for Graduate Education Research

Adam J. Blanford, Editor

Jason B. Scott, Editorial Assistant

Copyright © 2017 Graduate Teacher Program, University of Colorado Boulder
Teaching-as-Research Projects at CU Boulder
To the next generation of STEM educators.
# Table of Contents

**Chapter One**  
Introduction................................................................................................................................................1  
*Adam J. Blanford*

**Chapter Two**  
Developing Teaching-as-Research Projects at CU Boulder.................................................................12  
*Jason B. Scott and Adam J. Blanford*

**Chapter Three**  
Assessment of Learning in a Graduate-Level Pedagogy Course.........................................................22  
*W. F. Preston Cumming*

**Chapter Four**  
Belonging and Ability Uncertainty in LGBTQ+ General Chemistry Students...............................45  
*Jay Kroll*

**Chapter Five**  
How Does the Implementation of Case Studies in a Second-Semester General Chemistry Course for Majors Affect Student Learning? .................................................................57  
*Anna Curtis*

**Chapter Six**  
Quantifying the Impact of a Physics Co-Seminar.................................................................................68  
*Daniel C. Cole*

**Chapter Seven**  
Student Feedback on Different Activity Formats Implemented in Upper-Division, Undergraduate Neurophysiology Lab Course at the University of Colorado Boulder.................................................90  
*Diba Mani and Adam Blanford*

**Chapter Eight**  
Student Feedback on Different Activity Formats Implemented in Upper-Division Undergraduate Neurophysiology Lab Course at the University of Colorado Boulder Following Pilot Study Modifications from the Previous Semester.........................................................................................110  
*Diba Mani*

**Chapter Nine**  
How Does the Extent of Active Learning Present in an IPHY Lab Course Affect the Information Retention and Learning Gain of Statistical Concepts among Undergraduate Students?...........141  
*Christopher Stamper and Adam J. Blanford*

**Chapter Ten**  
Characterizing the Accessibility Climate in Geology.............................................................................156  
*Emily Fairfax and Megan R. M. Brown*
Chapter Eleven
Effects of Learning-By-Teaching in a Flipped Classroom

Victoria Li, Wendy Bailey, Ray Littlejohn
Chapter One

Introduction

Adam J. Blanford, PhD
TIGER Coordinator, Graduate Teacher Program

This book explores a sampling of Teaching-as-Research (TAR) projects performed by fellows of the Teaching Institute for Graduate Education Research (TIGER), a unit of the University of Colorado Boulder Graduate Teacher Program and a node in the Center for the Integration of Research, Teaching, and Learning (CIRTL) Network. As a member of CIRTL, TIGER’s mission is to improve the quality of graduate teaching and learning in Science, Technology, Engineering, and Mathematics (STEM) fields. We do this by showing graduate students how to improve awareness of their teaching, the quality of their interactions with students, and how they can systematically assess and improve their teaching.

CIRTL addresses a prevailing problem in higher education; namely, that research and publication, although vital components of success in academia, receive much greater emphasis in graduate professional development than teaching. Teaching is regarded as a secondary endeavor, deferred until later when the graduate student is about to become a member of the professoriate. CIRTL uses its three “core ideas” (Learning Communities, Learning-through-Diversity, and Teaching-as-Research) to approach the classroom environment systematically to understand the dynamics in a particular class and to develop a broad array of learning tools to facilitate deeper learning and connectedness among students. The core ideas also enhance the student’s ability to diversity their instructional methods.

The CIRTL Network is constantly growing and changing as its constituents develop deeper understanding of the core ideas, learn more about how to develop and sustain TAR projects, and make changes in response to shifting attitudes in higher education and society at large. In ten
years, CIRTL has grown from just six institutions to forty-three, which shows a growing recognition of the need for better STEM graduate teaching preparation. These expansions add to the Network’s diversity, as institutions with all manner of missions and organizations come together in support of this endeavor, each interpreting and implementing the CIRTL mandate and core ideas in their own way. As one of the oldest constituents in the CIRTL Network, TIGER has considerable experience working with its core ideas and creating TAR projects that contribute to STEM graduate development, all while enhancing the quality of undergraduate learning.

In this chapter, I provide an overview of CIRTL, TIGER, conducting TAR at CU Boulder, and the TAR projects represented in this volume. I begin with a description of the CIRTL Network and the “core ideas” that inform what we do as a network to further its mission. Next, I discuss TIGER, including its constituent programs, the ways TIGER has changed, and how we execute TAR at our institution. I also briefly describe the ten TAR projects that make up this volume. In doing so, this chapter will show some of the work TIGER has accomplished while contextualizing the environment in which the contributors to this volume worked on their TAR projects.

**The CIRTL Network: An Overview**

The CIRTL Network is currently an association of forty-three institutions across the United States and Canada dedicated to improving graduate teaching and undergraduate learning in STEM. The network was founded to address a need for better STEM educators in higher education, as well as a lack of preparation for graduate students looking to enter the professoriate
but without the requisite experience to teach in higher education or assess their teaching along the way.

The first iteration of the CIRTL Network formed in 2006 with six institutions: University of Wisconsin Madison, Vanderbilt University, Howard University, Texas A&M, Michigan State University, and CU Boulder. A growing realization of the need for better STEM educators precipitated the network’s formation. Eighty percent of STEM PhDs are produced by one hundred institutions in the United States (i.e., the “Research University/Very High” classification (Carnegie 2017) (Pfund et al. 2006). The focus on research, publication, and grants is instilled in graduate students from the very beginning, creating a situation in which teaching and learning become secondary. Graduate students emerge with their graduate degrees and impressive research credentials, but when they take professorships they are often ill-equipped to teach in the classroom. This is not to say they cannot improve with time, but the learning curve is steep because it takes place when they are already junior faculty on the tenure track, with potentially less time and room for error.

It is clear that more intense focus on graduate preparation is needed because of the importance of graduate students as disciplinary ambassadors and the inherent benefits teaching has for graduate development. Research suggests that undergraduates are twice as likely to inspire undergraduates to switch majors compared to tenured or tenure-track faculty (Bettinger et al. 2016; Flaherty 2016), because graduate students are viewed as more accessible and they have closer overall contact with undergraduates. Moreover, studies show that graduate students actually improve their research skills through teaching because examination and presentation of the basic material promotes metacognitive processes that contribute to the graduate student’s deeper understanding of his or her discipline (Feldon et al. 2011).
In this vein, CIRTL harnesses the synergy of teaching and research to delve deeper into what Boyer (1990) calls the Scholarship of Teaching and Learning, by using three core ideas to guide STEM graduate teachers. Learning Communities involve creating an open, equitable environment in the classroom in which all members of the community invest in the group to attain learning goals (CIRTL 2017). Learning communities promote collaborative learning and connectedness and inclusivity among all participants (faculty and students). In addition, learning communities emphasize the functional connections between the material and its utility in the real world.

Learning-through-Diversity emphasizes understanding the many types of diversity that exist in the classroom and recognizing them as potential tools to enhance the learning experience. Students have many and varied backgrounds, carry diverse views on a range of subjects, and each comes with a range of skills and experiences that may be brought to bear in the classroom. Interdisciplinary teams are actually more productive in the generation of knowledge because they approach problems from completely different angles. They also encourage greater communication between team members about their reasoning and ideas (Wuchty et al. 2007).

Teaching-as-Research is the idea that teaching is worthy of scholarly research and analysis. Researchers ask small-scale questions about their teaching, consult the literature, formulate hypotheses, gather and analyze data, and draw conclusions. TAR is unique because it is both a concept and a practice that embodies the connections between teaching and research. These core ideas help the student understand his or her teaching, create a greater awareness of the classroom environment, push them to reflect on their teaching approach, and make changes accordingly. The three ideas inform the work that we do through TIGER at CU Boulder.
The Teaching Institute for Graduate Education Research (TIGER)

TIGER is CU Boulder’s node in the CIRTL Network, tasked with promoting the network and its core ideas, as well as providing graduate students with the tools necessary to improve their teaching. It is also a unit of the CU Boulder Graduate Teacher Program, a teaching and professional development program for graduate students that offers trainings, consultations, and certifications to graduate teachers and future faculty. Because TIGER is able to build on the infrastructure of the GTP, this arrangement has been very beneficial to TIGER’s growth and sustainability. Since its founding in 2006, the program has experienced considerable change with the creation of new programs and technologies. In addition, we have made changes in response to lessons learned while during development. In this section, I briefly describe TIGER, including its pedagogical development, workshops and trainings, mentoring, and creation of TAR projects.

One of TIGER’s earliest projects was the Design and Development of College Pedagogy (TIGER DAD) project, a year-long sequence in which advanced graduate students worked with TIGER staff to develop a pedagogy course specific to their discipline. Graduate students worked with the GTP Director and TIGER staff to create a comprehensive syllabus, course lessons, reading lists, and class activities for use in a for-credit pedagogy course taught in the department. TIGER DAD produced three pedagogy courses, one in the Department of Aerospace Engineering Sciences (never taught), one in the Department of Atmospheric and Oceanic Sciences (taught once), and a cross-listed course in the Departments of Geography, Environmental Studies, and Ecology and Evolutionary Biology. This last course is still taught each year as a one-credit class in these departments. TIGER DAD was discontinued in 2012 due

1 Not all acronyms are gems…
to a lack of funding for graduate students and because the program only works if the department is willing to let the graduate student teach, or if a faculty member can be found to act as the instructor of record. Though efforts are being made to revive this program, the TIGER staff are focusing their efforts on current offerings.

TIGER hosts a number of in-person events like workshops and training intensives that cover pedagogy, classroom management, diversity, and academic retention. The TIGER workshops (TIGER 1: Workshops on STEM Teaching; TIGER 2: Workshops for Research Assistants and Postdocs) are among our most successful events, with approximately one thousand attendees over the last ten years. TIGER 1 workshops focus on STEM teaching and learning, with presentations by GTP staff, TIGER staff, and faculty on teaching techniques, classroom management, assessment, and the CIRTL core ideas. TIGER 2 workshops address aspects of professional development, with presentations of job searching and interviews, writing (e.g., proposals, teaching statements), and team management. We encourage attendance at TIGER workshops for STEM graduate students, especially TAR fellows, because of the useful techniques presented by experienced faculty, staff, and students; however, this is not always possible due to very busy research schedules and commitments.

In addition to workshops, TIGER also hosts three multi-day training events during the summer that cover topics of diversity, retention, and assessment. TIGER Diversity on Campus (TIGER DOC) started in 2013 to train STEM graduate students how to recognize classroom diversity, create inclusive environments, mitigate contentious situations, and minimize microaggressions (GTP 2017). Over the years, this training has undergone several changes, most notably the inclusion of arts and humanities students, as well as expansion to include two other University of Colorado campuses (Denver and Anschutz medical campus).
TIGER Summer Teaching-as-Research for Post-docs in Engineering (STRIPE\textsuperscript{2}) is a week-long summer institute that trains post-docs and new faculty on pedagogy and assessment for STEM classrooms. Similarly, TIGER Research on Academic Retention (ROAR) is a three-day training event started in 2015 to train graduate students in classroom management and assessment specifically so they could develop TAR projects. Trainings cover pedagogical techniques (e.g., Bloom’s Taxonomy), classroom assessment techniques (e.g., minute papers), and the processes involved with conducting a TAR project through TIGER.

Underpinning nearly all activities are one-hour mentoring sessions that we refer to as “TIGER Teaches.” During these sessions, graduate students meet with members of the TIGER staff to discuss pedagogy, TAR project development, assessment, survey techniques, and writing. Talking about aspects of teaching and professional development helps the graduate student think through a problem so they feel better prepared to work on their respective projects, while also developing the types of habits they will need after they graduate.

One of the most essential elements of TIGER is the TAR project, because they are actionable and most closely connected to graduate student teaching development. CIRTL promotes TAR as the creation of small-scale research projects that investigate areas of improvement in graduate teaching, as well as the use of the scientific method and research skills for assessment. These projects are intended purely for the teacher’s own interest and improvement; thus, the data are used solely with the classroom and not intended for public dissemination, which saves time in project development and implementation.

In contrast, TIGER TAR fellows must obtain human subjects certification and submit their projects to the university’s Institutional Review Board (IRB) for approval before beginning the

\textsuperscript{2} …although occasionally there’s a winner.
These steps may add extra time, but they ensure the project has been appropriately scaled and structured, with a clearly defined research question, research design, and methods. Moreover, this ensures that projects have appropriate safeguards to protect the subject populations and that the researcher has taken the time to obtain informed consent so they may use the research for subsequent projects and publish if they choose. TAR projects vary in the types of questions asked and in the overall scope of the research, depending on student interests, availability of data sets or classrooms in which to pursue the question, and the overall scope. As TIGER Coordinator, my job is to determine whether a project as proposed is feasible given the time and resources involved. I help the student frame the research question and determine the specific research design and data collection techniques required to address it. In addition, I advise the student through all aspects of submission to the IRB for approval. The TAR development process is discussed in greater depth in Chapter Two.

In seven years (2010–2017), TIGER has supported the development and execution of forty-five TAR projects in over a dozen departments on the CU Boulder campus. The ten projects presented in this volume are a sample of what we have accomplished with our program, and highlight topics of project development, pedagogy, inclusion, learning communities, and assessment. In Chapter Two, Jason Scott and I discuss the process we use for creating TIGER TAR projects, including recruitment, proposal development, completion, and write-up.

In Chapter Three, Dr. Preston Cumming presents the results of a TAR project he conducted in 2011 while he was graduate student in Geography. He chose to focus on the assessment of graduate student learning in a cross-listed, graduate-level pedagogy course (see TIGER DAD above). Dr. Cumming discusses the results of the project and evaluates it in light of his
intervening experience with higher education. In this project, one can see the development of a teaching scholar and understand the types of procedures TIGER started with in the beginning.

In Chapter Four, Chemistry PhD student Jay Kroll discusses how the degree of “belonging” and “ability uncertainty” (Lewis and Hodges 2015) among LGBTQ+ students taking general chemistry courses can affect their interest in Chemistry as a major and whether they feel the Chemistry department is inclusive.

In Chapter Five, Chemistry PhD student Anna Curtis discusses the use of case studies to promote the development of critical thinking skills among majors and non-majors in general chemistry courses. In general, she found slightly greater development of critical thinking skills among majors than non-majors.

In Chapter Six, Dan Cole, a PhD student in Physics, discusses the quantification of a physics co-seminar among first-generation college students. He shows that students taking the co-seminar benefit from the interactions.

Diba Mani, a PhD candidate in Integrative Physiology, discusses a two-semester project in Chapters Seven and Eight about activity formats and student learning in an integrative physiology lab during the 2016-2017 academic year. In Chapter Seven, she discusses the results of student surveys on which integrative physiology laboratory activities are most meaningful, while in Chapter Eight she discusses the results of those changes made as a result of student feedback during the spring semester.

In Chapter Nine, Christopher Stamper and Adam Blanford use the Differentiated Overt Learning Activity (DOLA) framework (Menekse et al. 2013) to observe the quality of student learning as they utilize statistical procedures in a revamped Integrative Physiology laboratory
course. They show that the changes made to the lab to encourage interaction and group work serve to enrich the classroom environment.

In Chapter Ten, PhD students Emily Fairfax and Megan Brown discuss the “accessibility climate” within the department of Geology, both in terms of accessibility of laboratory courses and activities (e.g., field trips) and the extent of faculty and teaching assistant knowledge of accessibility accommodations. Their findings suggest that changes should be made to make the department’s offerings more accessible and inclusive.

Chapter Eleven is presented by Victoria Li and her advisors Wendy Bailey and Ray Littlejohn, who discuss the results of a “learn-by-teaching” lesson given in a flipped engineering classroom. Though their findings show no discernible change in learning gains between flipped/traditional classrooms, the degree of interaction between students and teachers is greatly enhanced and contributes to an improved classroom environment.

These projects highlight the breadth and depth of potential TAR projects and the types of questions students may pursue as a result. TAR represents a valuable contribution to graduate student teaching development that will help produce the next generation of highly effective STEM educators.
References Cited


Chapter Two

Developing Teaching-as-Research Projects at CU Boulder

Jason B. Scott ¹ and Adam J. Blanford, PhD ²

¹TIGER Liaison, Graduate Teacher Program, University of Colorado Boulder;
²TIGER Coordinator, Graduate Teacher Program, University of Colorado Boulder

Introduction

The Teaching Institute for Graduate Education Research (TIGER) is a unit of the University of Colorado Boulder Graduate Teacher Program, with a mission to improve graduate teaching and learning in STEM through: 1) Incorporation of the CIRTL “Core Ideas” of Learning-through-Diversity, Learning Communities, and Teaching-as-Research (CIRTL 2017); 2) Training graduate students in pedagogy, classroom management, and assessment through workshops, training intensives, and one-on-one mentoring sessions; and, 3) the implementation of Teaching-as-Research (TAR) projects to show graduate students how to use their own disciplinary background and experience to treat teaching as a scholarly research project.

TAR projects are a critical component of graduate teaching development, as well as a core element of what CIRTL does because they instill in STEM graduate students an awareness of the classroom teaching environment and its constituents while also illustrating that one’s teaching can and should be improved. As a node in the CIRTL Network, TIGER works with graduate students to develop and implement TAR projects in a wide variety of classroom environments. To date, over forty TAR projects in more than a dozen disciplines have been carried out by TAR fellows in the past seven years.

Because CIRTL is a developing community of institutions across the United States, Canada, and beyond, the authors feel it would be a worthwhile contribution to discuss the TAR project development process we use at CU Boulder and how others might be able to use or modify the
process to suit existing programs or contribute to the development of new TAR programs at new CIRTL Network institutions. We begin with a description of the institutional structure within which TIGER operates and how this helps us promote our program on the CU Boulder campus. We continue with a discussion of the TAR recruitment process, including the methods we have found most effective for selling and incentivizing the projects to graduate students. Following recruitment is the project development process, including the human subjects research requirement, project research design, and institutional review board (IRB) submission. Next is the project execution process and how we ensure students complete the project and their final completion and write-up. These steps have led to a successful TAR program that has been going strong for a number of years.

TIGER and the Institutional Structure

TIGER is a unit housed within the CU Boulder Graduate Teacher Program (GTP) which in turn reports to the Graduate School. The GTP is a teaching and professional development program created specifically to support graduate students from all CU Boulder departments become better teachers and to prepare them for the next stage of their careers through workshops, training intensives (fall, spring, summer), teaching consultations, and site visits to other institutions. The GTP also supports a “Lead Network,” a group of approximately fifty graduate students (referred to as “leads”) who are trained to act as consultants, mentors, and leaders in more than forty-five campus departments, colleges, and schools, as well as act as liaisons between those departments and the GTP. The focus on graduate students and the existing infrastructure offered by the GTP and the Lead Network are advantageous to promoting the program and are integral to how TIGER operates.
TIGER exists to promulgate the CIRTL Core Ideas and provide graduate students in STEM/Social Science and more recently the Arts and Humanities with the training necessary to be better scholars. As a unit, TIGER is composed of a full-time TIGER Coordinator who oversees the development of TAR projects, runs TIGER workshops, training events, and works directly with CIRTL Central to develop Cross-Network programming like CIRTL Reads. There is also a part-time TIGER Liaison who assists the Coordinator, and TIGER receives logistical support from the GTP Director, the Lead Coordinators who oversee the Lead Network, and the GTP staff.

Programmatically, TIGER relies on core activities like workshops, training intensives, mentoring sessions, and TAR projects to provide training and act as point of contact and recruitment among graduate students and postdocs to do projects. We run two workshops devoted to STEM teaching and professional development, such as TIGER 1: Workshops on STEM teaching, which covers STEM pedagogy, presentation, classroom management and assessment for STEM graduate students. The second workshop, TIGER 2: Workshops for Research Assistants, deals primarily with professional development like job interviews, grant writing, and team management. In addition, we run training intensives like Research on Academic Retention (ROAR), a yearly training that specifically focuses on Teaching-as-Research and TAR project development. Another example is Diversity on Campus (DOC), which trains graduate students to recognize and incorporate diversity into their classrooms and teaching techniques. In addition to the trainings, we conduct sessions known as “TIGER Teaches,” which consist of one-on-one or small-group meetings to discuss issues in pedagogy, writing, TAR project development, and even article submission. It is a more personalized means
of interaction that addresses student needs and is essential to developing clearly defined TAR projects.

TAR projects are intended to be small-scale, classroom-oriented research projects that allow students to explore aspects of teaching and learning and develop ways to improve. These projects are by nature, social-science oriented, so students must have requisite training, including certification through the Collaborative Institutional Training Initiative (CITI) by taking the “Social and Behavioral Research Investigators and Key Personnel” module. This module acquaints students with the nature of human subjects research and what can occur if the research subjects are not protected (e.g., Stanford Prison Experiment, Nuremberg Trials). It also describes the levels of IRB review and what constitutes informed consent. Though none of our projects go to extremes, it is important to ensure that all TAR fellows understand the implications their research might hold and the potential impact on their research subjects.

One aspect of TIGER TAR projects is the requirement that all projects receive approval from the university IRB. CU Boulder has a vibrant institutional culture with a strong commitment to discipline-based education research, making it crucial for researchers to develop projects that are mindful of students as research subjects and have built-in protections for student welfare. Furthermore, modeling TAR projects with an eye toward IRB approval acquaints the student researcher with a clear structure for TAR project design. It is with this that we proceed to the discussion of the steps in the TAR projects.
Steps in the TAR Project

We break steps in TAR project development down into recruitment, development, human subjects research submission, project execution, and write-up. This section discusses each step in the process.

Recruitment

We recruit TAR fellows through emails, workshops/trainings/talks, and word-of-mouth. The Lead Network is invaluable as a built-in point of contact with campus departments as the GTP Lead Coordinators sent emails out to each lead which are passed on the graduate students directly. In addition, we recruit potential fellows through our TIGER workshop series (see above), training events like ROAR which specifically recruit students to do TAR projects, and through departmental talks on TAR and how it can improve one’s teaching and learning. These modes of contact promote interest in TAR and encourage students to meet one-on-one.

At the initial meeting, the TIGER Coordinator and Liaison meet with the student to discuss the steps involved in carrying out a TAR project and begin to sketch out rough ideas for the sorts of projects that might interest the student. Students must have a faculty mentor to advise them on the disciplinary aspects of the project, although if the student is unable to find a suitable faculty mentor the TIGER Coordinator, as a member of the research faculty, is able to fill the role. They must also obtain certification from the Collaborative Institutional Training Initiative (CITI – citiprogram.org) to perform human subjects research. This is not just a requirement of TIGER; rather, all individuals submitting protocols to the university’s institutional review board must have certification before applying. This includes the faculty mentor and anyone on the protocol as well.
Graduate students may be reticent to undertake a TAR project due to existing scheduling or time commitments, so it is important during that initial meeting to discuss the benefits the project will have for the student. We incentivize students by offering $500 for a TAR project to compensate the student for their time. Additionally, we emphasize the potential professional benefits to them, including the improvement of their teaching skills, the establishment of a research program into the scholarship of teaching and learning, and the possibility of publishing the results of their research in a scholarly research journal. It is also good training for graduate students because it is a good way to rehearse the steps involved in carrying out their eventual dissertation research.

Development

After the initial meeting, we proceed with the TAR project development through a series of TIGER Teaches sessions. We work with TAR fellows to ensure the project is both interesting and relevant to the fellow; the CIRTL Core Ideas are properly integrated into the project framework; and, the project is feasible given the time and resources involved. On the first point, fellows have to be passionate and engaged with the project or it becomes more likely they will not finish. We ask fellows to think carefully about the kinds of questions that resonate with them, about the things they wish to learn, and what can be done based on their existing research experience.

The Core Ideas provide a clear framework for conceptualizing the classroom environment and the individuals within; therefore, we work with the fellows to ensure they consider the aspects of classroom diversity and community that are most relevant. Using TAR, we help the
student mold the project to most benefit the individuals within while also presenting a realistic project design.

We also work with TAR fellows to ensure projects are both realistic and feasible. Graduate students are learning to conduct research, but there is a tendency to underestimate the time and resources needed to complete a project, or they design a project that is very ambitious that exceeds the time commitment for a project. For example, one student proposed an initial project idea that involved no less than seven data collection methods, including methods like in-class observations, videotape transcription, analysis, surveys, and interviews. Needless to say, the project was too large in its initial form so during TIGER Teaches sessions we worked to define the research question clearly and prioritize the data collection methods to get what we absolutely needed to address the question. Once the research question is developed, the student is given clearance to begin writing the research protocol.

Submitting the projects to the university IRB ensures that students are free and clear to publish the results of their work if they choose, but one of the other benefits to going this route is the IRB protocol template provides a clear structure for addressing all elements of project. The protocol asks students to define the research question and hypotheses, and provide a clear justification for why the research should be done using the academic literature as a basis. It also asks the student to define the populations being studied, how they will recruit research subjects, and what their procedures will involve for data collection. Students must also draft informed consent documents and specify what measures they will take to secure the safety and security of their research data.
Submission

After the TIGER Coordinator approves the research documents, the student receives permission to submit the project to the university institutional review board. All TAR projects are submitted under the “exempt” level of classification because it carries the lowest level of risk to any human subjects involved and it is the most expedient level of review. One additional level of insurance we have at CU Boulder to ensure that there is proper project oversight is that once a graduate student submits the project to our electronic administration system, the application is routed to the faculty advisor listed on the protocol. The advisor is required to sign off that they have reviewed the documents and are aware of the project or else it will not go forward. Approval generally takes a week, although we have received approvals in as little as twenty-two hours and after as much as six weeks if revisions are requested. Once approval is given, the student is allowed to begin the research project.

Execution

As the principal investigator for the research project, it is the student’s responsibility to follow the procedures listed in the protocol. The TIGER Coordinator is there to provide guidance and oversight, but since this is a professional development opportunity for the student the Coordinator does not exert authority over the student unless absolutely necessary. There are regular check-ins with the TAR fellow, but beyond that the fellow is working independently to ensure the success of the project.
Completion

Classroom research is exciting, compelling, and largely uncertain because the TAR fellow cannot always control all of the variables needed for project success. A brilliantly designed project will fail if the students are not interested in participating, or if the instructor forgets to allocate sufficient time for the project data collection or to email a survey link out to the class. This can lead to less-than-optimal outcomes.

If a project is not successful, meaning there aren’t enough data points to conduct analysis or the project couldn’t be carried out as intended, we help the student as best we can by analyzing all aspects of the project to look for weak points. Was the problem in the recruitment of research subjects, the methods used? Was the theory inadequate to explain the research problem? What could we change if we ran the project a second time?

We also ask the student to reflect on the skills and experiences they had because even “failures” can lead to positive outcomes. They may have been able to use a new analytical technique or theory that is of interest, or working with human subjects might provide a bridge to other research they are conducting.

Write-up

At the conclusion of the TAR project, we ask fellows to write a report on their project, findings, and conclusions. These reports are modeled after articles in disciplinary educational journals, which provides a sort of first draft of a potential research article that the student may pursue if he or she chooses to. The TIGER Coordinator works with the students to ensure the reports are high-quality and ready for presentation.
Conclusions

TAR projects are intended to teach students how to conduct research on their own teaching, and in so doing increase their awareness of the classroom environment and the students they teach. These projects also show that graduate teachers can and should refine their teaching methods.

We structure TAR projects to follow our existing professional development model.

References Cited


Chapter Three

Assessment of Learning in a Graduate Level Pedagogy Course

W.F. Preston Cumming, PhD
Lead Coordinator, Graduate Teacher Program, CU Boulder

Introduction
This chapter discusses the results of a Teaching-as-Research (TAR) project in which I assessed learning gains among graduate students in a graduate-level pedagogy course during the 2011 fall semester. My contribution to this volume differs from other chapters in this volume because a) I chose to study graduate students; b) I studied learning gains in a graduate-level pedagogy course; and, c) I did this project six years ago, when I was a graduate student in the Department of Geography and a member of the GTP Lead Network (see Chapter One). The intention was to assess the effectiveness of the course at teaching graduate students how to teach in disciplinary contexts.

Pedagogy courses are designed to teach students how to teach within their own disciplines by describing general pedagogical tools and techniques, as well as including more discipline-specific approaches to facilitate improved student learning. These courses benefit from studies of learning and learning gains in undergraduate classrooms which explore effective metacognitive strategies, gauging the depth of student engagement, and learning in context of both formative and summative assessments (Cook et al. 2013; Bridges et al. 2002; Brookheart 2001; Koljatic and Kuh 2000) because the findings allow pedagogy instructors to review and make improvements. However, the study of graduate student learning is relatively unexplored, especially where it pertains to intersections with pedagogy and professional development.

Improving the quality of graduate student learning, especially where it pertains to pedagogy and teaching, is essential given the inherent challenges new generations of graduate
students face. For example, graduate students today find themselves working with an increasingly diverse university population in both ethnicity and academic backgrounds (Austin 2002; Cody and Hagerman 1997), which requires pedagogy that is timely, relevant, and calibrated to maximize learning among graduate students so they utilize it to pass on their knowledge to undergraduates. To succeed in future professions or academia, graduate students will need opportunities to engage with the scholarship of teaching in a more diverse climate (Smith et al. 2013; Trautmann and Krasny 2006; Golde and Dore 2001). This will not only lead to them being better educators, but has also been shown to increase research skills in graduate students (Feldon et al. 2011).

The objective of this study was to gain immediate feedback on the course’s ability to educate graduate students about pedagogical practices for effective teaching and learning in the classroom as well as evaluate the increase in the students’ in-class effectiveness after taking the course. In the following sections, I discuss the theoretical framework for the study and the use of pre-test/post-test assessments as tools to gauge learning and affect course adjustment. Next, I discuss the particulars of the course, followed by the research design and a discussion of the analysis, which show modest graduate student learning gains over the course of the semester and possible areas for improvement. I conclude with a discussion about what this study means for graduate student pedagogy training and what I learned from doing a TAR project.

**Theoretical Framework**

Pedagogy courses are designed to teach students how to teach within their own and related disciplines. In the pedagogy course described here students were asked to take pre- and post-assessments composed of ten different questions to evaluate participant knowledge and
confidence in their abilities to institute pedagogical methods into their classes. The questionnaire was given the first day of the course and then the last day of the course to assess the overall learning of the students of the methods presented to them.

The idea behind classroom assessment involves student and teachers in the continuous monitoring of students’ learning (Jacoby et al. 2014; Bell and Cowie 2001; Black 1998; Sadler 1998). It provides faculty with feedback about their effectiveness as teachers, and it gives students a measure of their progress as learners. Most important, because classroom assessments are created, administered, and analyzed by teachers themselves on questions of teaching and learning that are important to them, the likelihood that instructors will apply the results of the assessment to their own teaching is greatly enhanced. An excellent approach to evaluating increases in student understanding and knowledge gain is a pre- and post-assessment. Pre- and post-assessment designs are widely used in educational research, primarily for comparing groups and/or measuring change resulting from teaching technique changes (Gosselin and Macklem-Hurst 2002; Zimmerman and Williams 1998; Dugard and Todman 1995). In educational research, change is commonly measured in such variables as meeting of course goals and objectives, student understanding of processes, and student ability to obtain higher levels of thinking. The measurement of change provides a vehicle for assessing the impact of pedagogical strategies and processes, as well as the effects of specific methodologies and interventions.

**Course description**

The pedagogy course studied in this TAR project is cross-listed in the departments of Environmental Studies, Ecology and Evolutionary Biology, and Geography and has been taught for seven years (2010-2017). It is typically offered for one credit-hour and meets once a week.
An Environmental Studies graduate student developed the course in 2010 with mentorship from the Director of the Graduate Teacher Program. The student facilitated the course during the fall semester in Environmental Studies, although graduate students from all campus departments were invited to take part. Following its initial success, the course was crossed listed in Ecology and Evolutionary Biology and Geography. It is now common for the lead graduate teachers from each department to facilitate the course because of the existing training and experience in pedagogy.

The course was designed to provide students with the knowledge and skills required for implementing new techniques, improving their teaching skills and designing effective class materials. Teaching strategies for the course included in-class and online discussions, provided readings, in-class activities, group work, course unit preparation and lectures. The class requirements consisted of participation in class discussions both in class and in an online WIKI format, active engagement in activities and demonstrations and a group project in which students produced an interdisciplinary course unit surrounding a topic of their choosing which was then presented to the class as a capstone project. The interdisciplinary focus of the pedagogy course has given new insight into teaching for all three departments involved.

This project took place in the fall of 2011 with a study population of 18 students (6 male and 12 female) that met once a week for the 1-credit course. Students ranged in classroom experience from 1-7 semesters in the classroom to that point. The mix of students enrolled also varied in terms of their current graduate program standing, with 9 being in an M.S. program and 9 a Ph.D. program. Fifteen of the 18 students expressed interest in future academic careers. It is interesting to note that the few remaining students (3) indicated that teaching was not in their...
career plans, but expressed that the knowledge that they will gain could only benefit them in the job place.

**Research design and methods**

For this study, we used a simultaneous mixed method design, where both the quantitative and qualitative data were collected and analyzed at the same time (Teddie and Teddie 1998). For the quantitative data, a paired *t*-test was run to compare the repeated measure (i.e. the pre- and post-assessment) on the same group to determine whether the means differ. The responses to open-ended questions are the most raw and unaffected parts of this and any survey analysis. The questions are completely unaided, and respondents can say or write anything that comes to mind. They are not limited to the selecting choices or guided in their response. With open-ended questions, you get a true sense for how the respondents feel and how to most effectively adjust course material and design for future classes.

I utilized a pre- and post-assessment to quantitatively and qualitatively analyze the gains in understanding and implementation within each type of pedagogical method as well as comparisons between types of methods. Evaluation through pre- and post-assessment assessment is used in many disciplinary contexts (Crisostomo 2010; Enos 2010; Robelia et al. 2010; Beers 2005), this format is normally selected to measure student-learning outcomes in each course type. The pre- and post-assessment format allows instructors to determine which learning outcomes they wish to test for and develop a testing instrument accordingly (University of Wisconsin-Madison 2009; Askins and Rossi 2005). This type of assessment assumes that improvement occurs between the pre- and post-assessment and that this improvement can be attributed to learning that takes place over the time-period of instruction, whether from in-class
instruction or out of class learning (Askins and Rossi 2005). These studies generally utilize a single pedagogical method that may not be the norm within that discipline to assess whether student learning increases through various methodologies.

One additional source in assessing the improvements made by students in this course was to use Faculty Course Questionnaire (FCQ) data freely obtained on the University of Colorado Boulder’s Planning, Budget, and Analysis program (available at www.colorado.edu/pba/fcq). Departmental, Division, and Campus wide averages for several categories (course and instructor overall, instructor effectiveness, and how much students learned) were used as proxies for student learning and provided an effective comparison with courses taught by graduate students that have not taken this course. As no department that this course is currently open to provides a similar course, the comparisons provide a useful viewpoint.

Results and Discussion
The following section discusses the results of the pre-/post-assessments conducted during the fall 2011 pedagogy course. The assessments consisted of both 1-5 Likert-scale (1=disagree, 5=agree) and open-ended questions about student knowledge and articulation of pedagogical techniques covered during the course of the fall semester. For each question, I discuss the statistics of the Likert-scale data and the summary of free responses. For every question, there was a highly significant (p<0.01) learning gain by each student (Table 1). The only exception was question 1 (“Could you please name and describe any learning styles that you are aware of?”), in which the increase of 0.44 (p value of 0.14) was not significantly different. This is most likely due to the highest pre-evaluation value of 3.89 compared to all other questions. The
results indicate significant learning gains in the quality of graduate student knowledge about pedagogy.

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Pre-mean</td>
<td>3.89</td>
<td>2.89</td>
<td>2.83</td>
<td>2.25</td>
<td>1.89</td>
<td>2.28</td>
<td>2.72</td>
<td>1.41</td>
<td>2.06</td>
<td>2.00</td>
</tr>
<tr>
<td>Post-mean</td>
<td>4.33</td>
<td>4.78</td>
<td>4.56</td>
<td>4.61</td>
<td>3.89</td>
<td>3.78</td>
<td>4.39</td>
<td>3.33</td>
<td>3.75</td>
<td>4.25</td>
</tr>
<tr>
<td>Difference</td>
<td>0.44</td>
<td>1.89</td>
<td>1.72</td>
<td>2.36</td>
<td>2.00</td>
<td>1.50</td>
<td>1.67</td>
<td>1.92</td>
<td>1.69</td>
<td>2.25</td>
</tr>
<tr>
<td>Increase</td>
<td>9%</td>
<td>38%</td>
<td>35%</td>
<td>47%</td>
<td>40%</td>
<td>30%</td>
<td>33%</td>
<td>38%</td>
<td>34%</td>
<td>45%</td>
</tr>
<tr>
<td>p-value</td>
<td>0.14</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 1. Likert Scale answers for pre- and post-assessment data.

**Likert Scale Results**

The pre- and post-assessment questionnaire (see Appendix 1) was designed to determine the level of student learning in the course. Each student was asked to rank their current knowledge of and/or ability to implement a certain pedagogical method in their classroom and were then given the opportunity to expand on that answer in an open-ended format. These questions used a 1-5 Likert Scale (1=disagree, 2=somewhat disagree, 3=neither agree nor disagree, 4=somewhat agree, 5=agree).

**Question 1: I can describe several learning styles that are common among students.**

Eighteen students answered this question on the pre- and post-assessments. The mean for the pre-assessment was 3.89, which is slightly below “somewhat agree” on the Likert Scale, while the post-assessment mean was 4.33. This indicates a modest 9% increase in agreement.
Question 1A: Could you please name and describe any learning styles that you are aware of?

Pre-assessment answered on learning style varied widely. The most commonly discussed learning styles were visual, auditory, and kinesthetic had the most responses with 10, 11, and 10, respectively. Active (4), passive (3), concrete (1), abstract (3), and logical (1) had very few occurrences in the discussions. The lecture on learning styles was coordinated with a discussion of various styles, but the lecture itself had an emphasis on the Kolb learning styles inventory (Kolb 1984). Kolb saw learning style as the unique learning method presented by the learner during the learning process and situation (Wang et al. 2006). It was interesting to note that the Kolb (1984) styles were discussed four times in the pre-evaluation. This was most likely due to the Graduate Teacher Program’s influence and attendance throughout the orientation week leading up to classes during the fall semester as well as the range of experience of the graduate students in the course.

The post-assessment returned a similar outcome, but with fewer categories of responses. Visual, auditory, and kinesthetic learning all were discussed on 12 of the responses, while Kolb, interestingly was only discussed on three. The number of responses that did not recall was also three. Having the Kolb learning styles discussed in fewer responses may indicate that there needs to be more direct reference and maybe even a requirement of all students to take the inventory to have it better outlined for them. Although the directive of the class period may not have been met, the students did seem to respond well to various learning style discussions.

Question 2: I can design a lesson plan that encourages each of these learning styles.
Pre-assessment responses totaled 2.89, which is near “neither agree nor disagree” on the Likert scale. The post-assessment mean was 4.78, indicating a significant jump in agreement on their ability to design lesson plans. There was no free-response question.

**Question 3: I know how to lead several different activities that engage all students in a classroom discussion.**

The pre-assessment mean was 2.83, while the post-assessment mean was 4.56. This is a 35% increase in agreement over time.

**Question 3A: Could you please name and describe any activities that would engage students in classroom discussions?**

Answers to this question varied widely on both the pre- and post-assessments. Most of the responses on the pre-evaluation were based on group discussions (10 responses) and question and answer times during the class period (6 responses). However, there were numerous other responses including using lecture, videos, debates, games, and think-pair-share activities that all were discussed at least once within the answers. This class period focused on the think-pair-share and gallery walk activities with a specific discussion thread opened on the class wiki to include any other ideas for classroom discussion.

Within the post-assessment answers it is evident that the class objective was met with a significant increase in responses for think-pair-share (11), gallery walk (13), and discussion based activities (13) respectively. The diversity in answers significantly increased to now include question and answer time (2), debates (3), role playing (5), games (2), case studies (3), classroom research projects (1), presentations (1), and videos (1). This may be due to the class wiki posts.
for activities as well as the classroom discussion following the two activity demonstrations in class.

**Question 4:** I can use Bloom’s Taxonomy as a guide for writing quiz/test questions for each of Bloom’s Hierarchy.

Bloom’s Taxonomy of the Cognitive Domain (REF) is a pedagogical tool consisting of six categories – Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation – of increasing cognitive complexity and interaction with course material that teachers can use to develop course content that is attainable and relevant to the course objectives.

For the pre-assessment, the mean was 2.25, indicating that respondents were more in disagreement with the question premise. However, the post-assessment mean was 4.61, which suggests a marked improvement in their level of agreement for writing questions based on Bloom’s Taxonomy of the Cognitive Domain.

**Question 4A: Could you please write a question for each of Bloom’s hierarchy?**

One of the most significant increases in correct responses from the pre- to the post-assessment occurred on the Bloom’s Taxonomy (REF) question. Answers to the pre-evaluation were very low with knowledge (2), comprehension (2), application (3), analysis (3), synthesis (2), evaluation (1), and unsure (5) showing little or no understanding of the taxonomy or at least a lack of recall from attended Graduate Teacher Program workshops. The post-assessment significantly increased with knowledge (16), comprehension (9), application (9), analysis (12), synthesis (9), evaluation (7), and unsure (5). It is also important to note here that only one student could name all levels specifically.
Student responses to this question varied; however, every student with their notes in front of them could accurately identify and utilize all the taxonomic levels within their course units for the final project. Many of the students who replied partially or not at all were explicit in identifying this fact. It is also interesting to notice that most students could accurately identify and write a question for the basic level (knowledge), but the numbers of correct responses decreased as the definition and questions became more difficult to describe.

The class period devoted to this assessment technique was designed not only to discuss Bloom’s taxonomy, but also other modes of assessment. Students were given the opportunity to write questions and have them “graded” for correctness. Only a handful of students took this opportunity and it may be part of the reason for not completely grasping the concepts at every taxonomic level. However, the classroom discussion and wiki thread were both fruitful and highly useful to all students and instructors.

**Question 5:** I can implement a classroom activity that uses the Problem Orientation Framework for helping students analyze a problem.

The pre-assessment mean was 1.89, which suggests little or no familiarity with this framework since the mean score corresponds to “somewhat disagree.” Although the post-assessment shows a two-point increase in the mean, it is still slightly below “somewhat agree,” which suggests that the students still do not feel that they can implement a meaningful classroom activity using the Problem Orientation Framework.
Question 5A: Describe how you would implement a classroom activity that uses the Problem Orientation as a framework for helping students analyze a problem.

The discussion on the Problem Orientation framework was designed to have students look at how instructors can approach an interdisciplinary course that is focused, whether one class period or throughout a unit, on solving a larger problem such as climate change or global health care using a four-step method that includes the goals of the discussion, current trends in the issue, conditions of both sides of the issue, projections of outcomes, alternatives to the outcomes. Many students were Environmental Studies graduate students and may have could answer this question prior to the course. However, there were no responses on the pre-evaluation indicating either a misunderstanding of the question or it was not posed in a way to have them comprehend the meaning.

After the course, there was a significant increase in student’s comprehension of the question and their ability to describe how they would approach the issue. Many students discussed the use of case studies (4), describing and utilizing both sides of an issue (4), using real world problems (4), but only a few indicated all the steps involved in the framework (2). Only one student indicated that they did not remember what the framework was or begin to describe how they would approach a problem oriented class.

Question 6: I can design a classroom activity that encourages students to work in learning communities to devise a solution to a current problem.

The pre-assessment mean is 2.28, corresponding to “somewhat disagree” on the Likert-scale. The post-assessment mean is 3.78, which shows an increase in agreement and possible learning
gains among graduate students, but not enough to show confidence at designing a classroom activity.

**Question 6A: Describe what a learning community is and how you would encourage students to solve a current problem using one.**

The pre-evaluation only had one student who accurately described what a learning community was. Whether this was based on prior knowledge or a good guess is undetermined. The Center for the Integration of Research, Teaching, and Learning (CIRTL) describes Teaching-as-Research (TAR) as “involving the deliberate, systematic, and reflective use of research methods to develop and implement teaching practices that advance the learning experiences and outcomes of students and teachers (CIRTL 2012). The treatment of a classroom as a laboratory for determining outcomes based on the scientific method is a natural setting for many STEM students. The determination of what students have learned in the classroom is at the core of teaching and, therefore, at the core of TAR. Instructors have an exceptional opportunity to use their classrooms as laboratories for the study of learning and through such study to develop a better understanding of the learning process and the impact of their teaching upon it.

By beginning with constructing a foundation of knowledge and progressing through the framework of TAR as follows:

- Learning foundational knowledge.
- Creating objectives for student learning.
- Developing a hypothesis for practices to achieve the learning objectives.
- Defining measures of success.
- Developing and implementing teaching practices within an experimental design.
- Collecting and analyzing data.
Reflecting, evaluating, and iterating (CIRTL 2012).

The current study follows these steps and begins to assess the learning ability of graduate students in a pedagogy course. The novelty of this research is defined by the ability of each of these students to potentially utilize the techniques in their own classrooms and potentially have TAR projects conducted in their own classrooms using each individual pedagogical technique they have now discovered. After the discussion of this basic CIRTL network pillar in the class period, 13 of the 18 students could define what a learning community was and 12 students were able to describe how they would use it to encourage students to solve a current problem. Many students either defined learning community or explained how they would approach it, with very few doing both.

**Question 7: I know how to integrate diverse issues and viewpoints into a problem-based curriculum.**

The pre-assessment mean for Question 7 is 2.72. The Post-assessment mean is 4.39, which suggests that students learned more about how to integrate diverse issues and could possibly integrate them into the curriculum.

**Question 7A: Could you describe how you would integrate diverse issues and viewpoints into a problem-based curriculum by giving one example of a diverse issue?**

The pre-assessment answers to this question varied with only five total responses including energy use (1), pine beetle infestation (1), open space issues (1), water rights (1), and politics (1). The post-assessment showed a significant increase in the diversity and response rate including
climate change (5), health care (3), evolution (2), ranching (1), mining (1), developing country’s rights (1), and population control (1).

**Question 8:** I know how to practice Teaching-as-Research (TAR) as a method of assessing student learning.

Question Eight’s pre-assessment mean score was 1.41, which suggests almost no familiarity with the TAR concept. The post-assessment mean score was 3.33. These data indicate modest gains, but students likely do not feel they are capable of practicing TAR on their own.

**Question 8A: Could you describe what TAR is and how it is used as an assessment tool?**

The pre-assessment results for this question had no responses. This indicated that students were not exposed to this terminology or topic. Utilizing this discussion as a second-class period for assessment showed in the post-assessment with increases in the definition of TAR (7), the use of this pre- and post-assessment as a method for a TAR (8), the discussion of scientific research (3), and using TAR as an assessment tool (7) all being discussed. Although the specifics of TAR were not given as much time during the class period as would have been preferred, students did seem to grasp the concept and could see how the scientific method can be used as a means for assessment in a classroom.

**Question 9:** I understand the differences between classroom assessment and course assessment.

The question nine pre-assessment mean was 2.06. The post-assessment mean was 3.75.
Question 9A: Could you describe what the difference between classroom and course assessment is?

The answers on the pre-evaluation showed that only one student could accurately define what classroom assessment is and only one (different) student could define course assessment. Through the classroom discussion, the WIKI posts, and the development of their individual projects, those numbers significantly increased to 11 students in each category (many of which were the same) in the post-assessment.

Question 10: I feel that I can design a problem-based course that uses an interdisciplinary framework to highlight themes that are common among problem categories.

The pre-assessment mean was 2.00 (“somewhat disagree”), while the post-assessment mean was 4.25. This two-point jump shows an increase in student confidence at using an interdisciplinary framework.

Question 10A: Could you outline (bullet points) how you would go about designing a problem-based course that uses an interdisciplinary framework to highlight themes that are common among problem categories?

Within the pre-evaluation only one student attempted to outline how they might go about designing an interdisciplinary course. This number increased to 15 students being able to outline a course accurately with only three students failing to answer the question for the post-assessment.

The increase in student responses to the open-ended questions indicates the ability of graduate students to learn how to become more effective teachers in their classrooms. Many of
these students were either first or second time instructors and were potentially very nervous about either being in front of students for the first time or wanting to provide the best experience they could for the undergraduate students. This may indicate why these students enrolled in this type of course. The answers and explanations provided in this study will help direct the future of this pedagogy course in general and will provide information necessary to help other pedagogy courses based on this concept in other departments.

Finally, by comparing the graduate student’s faculty course questionnaire results from courses taught prior to taking the course and afterwards we see how effective such a course can be. Before taking the course, all students rated lower than the department, division, and campus overall in course overall, instructor overall, effectiveness, and only slightly higher than the department specifically in how much students learned in the course (Table 2). However, after taking the course, the average student results for all four categories show significant improvement and became higher than department, division, and campus averages (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Course</th>
<th>Instructor</th>
<th>Instructor Effectiveness</th>
<th>How Much Learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Pre</td>
<td>4.23</td>
<td>4.89</td>
<td>4.50</td>
<td>4.36</td>
</tr>
<tr>
<td>Post</td>
<td>4.57</td>
<td>5.22</td>
<td>4.85</td>
<td>4.59</td>
</tr>
<tr>
<td>Difference</td>
<td>0.34</td>
<td>0.33</td>
<td>0.35</td>
<td>0.23</td>
</tr>
<tr>
<td>Increase</td>
<td>5.6%</td>
<td>5.5%</td>
<td>5.8%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Department</td>
<td>4.40</td>
<td>5.10</td>
<td>4.65</td>
<td>4.35</td>
</tr>
<tr>
<td>Division</td>
<td>4.43</td>
<td>4.98</td>
<td>4.53</td>
<td>4.48</td>
</tr>
<tr>
<td>Campus</td>
<td>4.48</td>
<td>5.00</td>
<td>4.58</td>
<td>4.50</td>
</tr>
</tbody>
</table>

*Table 2 Faculty Course Questionnaire (FCQ) responses.*

Several factors may be involved in this increase including increased experience in the classroom and comfort with the material if in similar courses. However, it may also be attributed to the increased resources provided through a course designed to improve the graduate student’s ability to teach.
This project represented an attempt to move beyond the usual research on undergraduate education into independent assessment of student learning that can be translated directly into performance in the classroom. The findings show that students could take what they were learning about instruction and apply it to real world situations in their own classrooms. The overall instructional methodology grounded in the utilization of the same techniques the instructors were trying to teach the students, was responsible for engaging the students in active learning, creation of learning communities, and the increased awareness of learning through diversity.

Conclusions
The results of this study show that there may be many improvements made to the evaluation and in turn the course structure the further improvement of this and all pedagogy courses needs to be a constant process. The University as well as the educational community in general will benefit from this research, as the results will supplement literature on the assessment of teaching and learning in graduate level courses. Results from this research are being used currently for the development of similar courses in disciplines where graduate students are intending on either becoming faculty at an institute of higher education or in leadership roles in industry, government, or non-governmental organizations where having a Master’s or PhD level education injects you directly into a role as an instructor.

My initial expectations for this project were lofty and unrealistic. I was positive that it would be a ground breaking study with at least 30 students (to get my sample size high enough for significant statistics) that would contribute a novel angle on the way we look at graduate student learning. I soon discovered, however, that this was no different than anyone else’s
dissertation research. All I could say was scale down, scale down, scale down and the enrollment numbers forced me to do just that.

By undertaking this project, I learned a lot about where my priorities in higher education lie. I knew there was a passion for teaching in me and that it would come first wherever I ended up. I learned more about mixed methods, the IRB process, and what it takes to publish in educational development, whether I have gotten there yet or not. Coming from a strictly physical science/quantitative background with less understanding of the social science/qualitative approaches necessary in much of educational research I knew it would be an exceptional learning experience. My experience with TAR back in 2011 consisted of what I had learned through my work as a departmental Lead for the Graduate Teacher Program (GTP) in AY 2010/11 and subsequently AY 2011/12. In 2012 I was fortunate enough to go on a network exchange funded through the GTP and the Center for the Integration of Research, Teaching and Learning to the Texas A&M University to give presentations on not only this research project, but also my own dissertation research in the Geography Department. Since the time of this research I have returned to the GTP as a Lead Coordinator and have been happy to help with reviewing several IRB protocols and now involved with one TAR project as a co-investigator. I have continued to enjoy an enhanced role in educational development and the scholarship of teaching and learning through this position.

In the future, I expect to continue to strike a balance between my love of teaching and understanding more about educational research (SoTL/TAR) and my own research in grassland ecology. Fortunately, my current position with the Graduate Teacher Program allows me to maintain this balance, although a bit more on the educational development side of the spectrum.
References Cited


Appendix 1

Questionnaire

Part 1. Please complete the following survey by circling your answer according to this scale:

1 – Disagree
2 – Somewhat disagree
3 – Neither agree nor disagree
4 – Somewhat agree
5 – Agree

1. I can describe several learning styles that are common among students.
   1 2 3 4 5

1a. If you agree with number one, could you please name and describe any learning styles that you are aware of.

2. I can design a lesson plan that encourages each of these learning styles.
   1 2 3 4

3. I know how to lead several different activities that engage all students in a classroom discussion.
   1 2 3 4 5

3a. If you agree with number three, could you please name and describe any activities that would engage students in classroom discussions.

4. I can use Bloom’s Taxonomy as a guide for writing quiz/test questions that target different levels of students’ understanding of a concept.
   1 2 3 4 5

4a. If you agree with number four, could you please write a question for each of Bloom’s hierarchy.

5. I can implement a classroom activity that uses the Problem Orientation as a framework for helping students analyze a problem.
   1 2 3 4 5

5a. If you agreed with number five, describe how you would achieve this?

6. I can design a classroom activity that encourages students to work in learning communities to devise a solution to a current problem.
   1 2 3 4 5

6a. If you agreed with number six, describe what a learning community is and how you would achieve this?
7. I know how to integrate diverse issues and viewpoints into a problem-based curriculum.

1  2  3  4  5

7a. If you agree with number seven, could you describe how you would achieve this goal by giving one example of a diverse issue.

8. I know how to practice Teaching as Research (TAR) as a method of assessing student learning.

1  2  3  4  5

8a. If you agree with number eight, could you describe what TAR is and how it is used as an assessment tool.

9. I understand the differences between classroom assessment and course assessment.

1  2  3  4  5

9a. If you agree with number nine, could you describe what the difference between classroom and course assessment is.

10. I feel that I can design a problem-based course that uses an interdisciplinary framework to highlight themes that are common among problem categories.

1  2  3  4  5

10a. If you agree with number ten, could you outline (bullet points) how you would go about designing this problem-based course.
Chapter Four

Belonging and Ability Uncertainty in LGBTQ+ General Chemistry Students

Jay. A. Kroll
Department of Chemistry and Biochemistry, CU Boulder

Introduction

This chapter discusses a study focused on social belonging and ability uncertainty in Lesbian, Gay, Bisexual, Transgender, Queer/Questioning Plus (LGBTQ+) students taking general chemistry at CU Boulder. Building a diverse group of people in STEM (Science, Technology, Engineering, and Mathematics) is important to the continued success of these disciplines because having a diversity of people brings a diversity of thought processes and ideas that lead to greater productivity and success (Ellison and Mullin 2010; Nielsen et al. 2017; Nelson and Rogers 2003; Woolley et al. 2010; Woolley et al. 2015). Even though there is great benefit to diversity, in general, women and racial/ethnic minorities are underrepresented in STEM fields (Allen-Ramdial and Campbell 2014; NSF 2017). Because of this, there has been a broad interest in research studying the lack of diversity in STEM fields.

However, these studies tend to focus only on women and underrepresented racial/ethnic minorities (URMs). This unfortunately, leads to a gap in research on other underrepresented groups such as gender and sexually diverse people. Therefore, the needs of these people are often not focused on when evaluating how to increase diversity in STEM fields and they are further left behind, even when there are attempts to broaden representation.

While there have been studies of sexualities in STEM (Patridge et al. 2014; Riley 2008; Cech and Waidzunas 2011; Atherton et al. 2016; Yoder and Mattheis 2016; Bilimoria and Steward 2009), they tend to focus on broad groups, such as LGBT people in engineering (Riley
2008; Cech and Waidzunas 2011), or to focus on groups later in the STEM pipeline (e.g. professionals, faculty, and graduate students) (Atherton et al. 2016; Yoder and Mattheis 2016; Bilimoria and Stewart 2009). There are a number of positive responses in these studies, however, as the “Queer in STEM” study notes, “positive experiences reported by the majority of participants are encouraging but may mask those of individuals who have already self-selected out of the STEM career pipeline” (Yoder and Mattheis 2016: 21). This is particularly important when evaluating the “leaky pipeline” that leads to careers in STEM because if we want to increase LGBTQ+ representation, we need to know where in the STEM career pipeline we are losing LGBTQ+ people. One of the potential places of interest to investigate is the undergraduate classroom.

While the majority of current studies focus on later career STEM professionals, two of the studies cited above included responses from undergraduate students that help to illuminate the environment experienced by LGBTQ+ college students in STEM. The American Physical Society (APS) conducted a broad survey that included APS members at all stages in their careers. The APS climate survey collected responses from LGBT physicists (term used in report) including 62 undergraduate students (19% of respondents) (Atherton et al. 2016). The authors note in their analysis that a significant fraction of respondents had witnessed or experienced exclusionary behavior including “sexual harassment, verbal harassment, homophobic comments, purposeful misidentification of gender, exclusion from study groups and social activities, LGBT stereotyping, and expectations of incompetence” (Atherton et al. 2016:7). Additionally over one third of respondents had considered leaving their workplace or school in the last year and that observing exclusionary behavior strongly correlated with this consideration (Atherton et al. 2016).
Cech and Waidzunas (2011) focused their study on undergraduate and graduate students in an engineering program at a research university in the Western United States. They found through interviews and focus groups that there were pervasive prejudicial cultural norms and that perceptions of competence particular to engineering can limit LGB students’ ability to succeed, relative to their heterosexual peers (note that the authors did not focus on gender diverse populations and only studied lesbian, gay, and bisexual identities). The authors report that students navigate this chilly and heteronormative atmosphere by employing three tactics: 1) “passing” as heterosexual, 2) “covering” or downplaying characteristics that are associated with LGB identities, and 3) developing expertise that would make themselves indispensable to others (Cech and Waidzunas 2011). These tactics result in an increased emotional and academic workload for these students.

These two studies together point towards a less accepting and exclusive environment in undergraduate STEM programs, though they are relatively limited and may not necessarily be generalizable to all of STEM. It was with this in mind, and an interest in understanding where we begin to lose students out of the STEM career pipeline, that I chose to develop a survey to look at the sense of belonging for LGBTQ+ identified students in general chemistry. General chemistry students tend to be first- or second-year college students at or near the beginning of the college career. Additionally, a wide range of students from STEM and non-STEM majors take general chemistry, providing a unique pool of students to survey. Additionally, none of the previous studies in the literature have focused specifically on experiences in the chemistry classroom. Therefore, a survey of general chemistry students would provide a new perspective from LGBTQ+ undergraduate STEM students.
Theoretical Background

A sense of belonging in any situation is important to successful outcomes. Feeling uncertain about one’s social belonging can lead to questions about if one fits in and can lead to interpreting incoming information and cues with a negative bias, particularly in underrepresented or minority groups (Walton and Cohen 2007). Previous work suggests that sense of belonging is a significant predictor of positive academic outcomes, especially for groups that have been historically marginalized and stereotyped who may be more sensitive to perceiving and reacting to threats to their belonging (Lewis and Hodges 2015; Walton and Cohen 2011; Good et al. 2012; Johnson et al. 2007). Hodges and Lewis (2015) show in their paper that academic belonging is comprised of two related components, social belonging and ability belonging. They developed and validated a domain-specific measure of ability uncertainty, the sense that one has (or does not have) the abilities necessary to succeed in a field of study. Additionally they show that ability uncertainty is significantly correlated with social belonging, a sense that one fits in in a social context, but is a unique predictor of intent for students to persist in their major (Lewis and Hodges 2015).

I used a modified version of the Lewis and Hodges Ability Uncertainty Survey (2015) to ask students in general chemistry about their ability uncertainty in the course. Lewis and Hodges measured social belonging using two sub-scales from the Belonging to Math Scale (Good et al. 2012). While preparing my survey, I could not gain access to this study and could not find any belonging studies where the sets of questions were available in the text of the paper. Therefore, I developed my own set of questions to ask about social belonging and community. Students were then asked to self-identify their gender, sexual orientation, and intent to continue taking chemistry courses. I am currently analyzing the data from my survey.
**Previous Experience and Expectations**

I participate in the Chemical Education Research (CER) reading group at the University of Colorado where we read education research publications, discuss how they may apply to chemistry department at CU, and present and collaborate on ongoing education studies in the chemistry department. I also attend the Discipline-Based Education Research (DBER) seminars. This was the first education research project that I planned and was the lead researcher on. When approaching the Teaching-as-Research program my expectations were to find out if the study I was interested in conducting was possible. Once I decided that the study was possible, I conducted a pilot study to develop a survey that I could use in future research studies. I hoped to find out how many LGBTQ+ students we have in a typical general chemistry course and if those students are experiencing the course in a way that is different than their cisgender heterosexual peers.

With the help of the TAR program, and especially Dr. Adam Blanford, I set up a pilot survey in the 2016 Summer Session general chemistry classes and followed up that survey in the much larger general chemistry course in the Fall Semester of 2016. This has led to a data set containing responses from 722 unique general chemistry students across a wide spectrum of genders and sexualities. To the best of my knowledge, this is the first time anyone has broadly surveyed students in a general science course about their sexual orientation. While the goal of the survey was to study LGBTQ+ students and the data from the survey is still being analyzed, there are some interesting initial results when looking at the class as a whole. As shown below in Figure 1, ability uncertainty has some correlation with the final grade achieved in the class in both the pre- and the post-survey. In general, the more certain a student was that they had the
skills to succeed, the higher the grade they achieved in the course. It is worth noting, that in the pre-survey, this effect levels off for grades lower than a C.

![Figure 1](image.png)

**Figure 1** - Plot of ability uncertainty against final grade achieved in the course for the pre- (blue squares) and the post- (red circles) survey.

In Figure 2, we see there is a much stronger correlation of ability uncertainty with the grade students *expect* to get in the course. This is particularly interesting because even at the beginning of the course students have an expectation for the grade they will receive and this is likely connected to how strongly they believe that they have the skills necessary to succeed in the course or not. These achievement expectations can greatly impact the amount of effort a student puts into the coursework and thus have a real impact on the final grade achieved in the course.
In addition to grades, ability uncertainty and social belonging both appear to have an impact on classroom behavior. Students who are more certain in their social belonging and ability uncertainty respond that they ask questions more often in class and are more comfortable expressing their ideas in class. More analysis is needed to determine if there are statistical differences in these responses and to see if there is an impact in these measures related to gender or sexuality.

**Lessons Learned**

In developing the survey there are a few things that I have learned. The first is that in order to get any significant set of responses to my survey, it was important to get buy in from the professors who were teaching the course. Without the help of Dr. Margaret Asirvatham and the enthusiastic support of Dr. Kathryn Plath I would not have been able to collect such an extensive data set. Because I was able to interest them in my project, I was able to offer students a small
amount of course credit to take the survey. In my summer pilot study students took the survey on a purely voluntary basis with no course credit given. I was able to get seventeen responses from the entire course of approximately one hundred and twenty students in Summer Session A and another seventeen responses from Summer Session B general chemistry course of similar size. Comparatively, I was able to get seven hundred and twenty-two responses from the course of approximately one thousand students when I could offer them a small amount of course credit that amounted to 0.50% of their final grade.

Additionally, I learned a number of important things about the logistics involved in doing an educational research project, including the importance of clear instructions and guidelines for participation. When offering students course credit, it is important to clearly outline what they need to do in order to receive the credit. I made an announcement at the beginning of the course and told students that if they did not want participate in the study they did not have to, but that they needed still fill out the consent form and respond that they did not wish to participate in order to receive the class credit. This same information was included an email sent out to all students enrolled in the course. Even with the announcement and email there was confusion among the students about what they needed to do. At the end of the course, I received several emails from students who were under eighteen and could not legally participate in the survey but did not fill out the consent form and indicate they were unable or unwilling to participate in the survey. Because they could not participate, they did not believe they needed to respond at all in order to get the course credit. Be clear and repetitive about what is required for students to receive credit for a study if you are offering course credit.

In future surveys I intend to place the questions asking students to give their name and student ID for credit before the consent form. This way I can be explicit that all students need to
click on the link and provide their information in order to receive the class credit, then they will be asked to participate in the survey on a voluntary basis. This will make it clearer for students under eighteen years of age or students who do not wish to participate in the survey that they need to still go to the survey link to receive the credit for the course.

Another lesson I learned from this project was that you should have a clear plan about how you intend to analyze your data. My initial study has been very explorative and I wasn’t certain how much data or what sort of response rate I would receive. I had a plan for my analysis. However, general chemistry students tend to be in their first and second year of college. This is a time when college students often begin exploring identities such as gender and sexuality. I had not anticipated how this might appear in my data. This led to complications in my data analysis, raising questions of how to classify some students’ sexuality and gender. This has illuminated a series of new and interesting research questions but has certainly slowed progress on the data analysis.

The final lesson I learned is to give myself time. Finishing the pilot study late in the summer of 2016 did not give me much time to make any changes to my survey or analyze the data before administering it to the fall 2016 students. As a result, there are some parts of the survey that can certainly be improved. In particular the survey was fairly long (45 questions total) and slightly repetitive. In future versions, now that I can be more certain that the questions are asking what I intended, I may be able to remove some of the more repetitive questions. Additionally, the questions about gender and sexuality only allowed students to select one option, which may have led to some confusion for students about how to properly respond about what their identity is (e.g. how will a student who identifies as a woman and also identifies as
transgender respond?). After having more time to look closely at responses, I can find better ways to word questions that may reduce this uncertainty for future respondents.

**Future Directions**

I intend to finish analyzing my data and publish the results in a journal such as the *Journal of Chemical Education*. Additionally, I intend to do a more in-depth validation of my survey questions with the help of Dr. Kathryn Plath and find ways to shorten the survey and make questions about gender and sexuality clearer and more flexible for future students. I also plan to continue with this study and administer an improved version of the survey to future general chemistry students. I hope to build a more comprehensive set of data about gender and sexuality of general chemistry students. At CU, there is a separate general chemistry course for students who are majoring in chemistry. As such, there are few responses to the original survey from students majoring in chemistry and my data set cannot tell me about the ratio of genders or sexual orientations of chemistry majors. While the current data set provides a wide range of responses from many STEM majors and tells us much about the perception of belonging in the general chemistry course, future inclusion of the majors’ course in the survey would provide greater insight into the effect of gender and sexual orientation on social belonging and ability uncertainty of chemistry students.
References Cited


Chapter Five
How does the implementation of case studies in a second semester general chemistry course for Majors affect student learning?

Anna Curtis
Department of Chemistry, CU Boulder

Introduction
Many students have difficulty succeeding in undergraduate General Chemistry courses. One important factor may be students’ inability to go beyond the simple memorization and basic understanding sufficient for success in high school to the analysis and application of knowledge necessary for success in a university. It is essential that students develop these abilities to foster their success in STEM degrees and to prepare them for future STEM careers. There is evidence that case studies may increase students’ abilities in these areas (Hutchinson 2000; Dori et al. 2003). However, none of these studies address the use of case studies in a second semester chemistry course where students tend to struggle the most with more complex chemical concepts such as the reactions of acids and bases. For this project, the author proposed that in-class case studies implemented in the second semester of a general chemistry course will increase students’ critical thinking skills.

Background
I was a third year PhD candidate with a background in physical chemistry before beginning my Teaching as Research (TAR) project and I was new to the educational research. My only background knowledge was from an hour-long TAR workshop presented by the Graduate Teacher Program (GTP) at CU Boulder. My TAR project was also a novel endeavor because I had no previous experience with educational research or any research in the social
sciences. However, I did have extensive experience as a teaching assistant and a desire to
improve teaching in my discipline. I taught 5 semesters of general chemistry recitations and
laboratories (2 per semester) before beginning my TAR project. The research question from my
TAR project arose from this experience.

Theoretical Framework

The goal of this research was to assess the effect of in-class case studies, in the form of
group activities, on students’ critical thinking skills in a first semester general chemistry course
for chemistry and biochemistry majors. Critical thinking skills are important for success in
university as well as in future career paths (Roksa and Arum 2012). In a survey of employer
priorities for college learning and student success conducted by the Association of American
Colleges and Universities (AACU), ninety-three percent of employers agree that “a candidate’s
demonstrated capacity to think critically, communicate clearly, and solve complex problems is
more important than their undergraduate major” (AACU 2013). Whether their career path
involves diagnosing a patient or reducing the amount of greenhouse gases in the atmosphere,
students will need to have good critical thinking skills to be successful.

Studies show that a significant amount of students in higher education fail to obtain a
measurable gain in critical thinking skills (Roksa and Arum 2011; Pascarella et al. 2011).
Undergraduate science courses tend to focus on knowledge instead of the scientific process
(Johnson and Pigliucci 2004). It has also been shown that higher education does not always
correlate with skepticism of pseudoscience and the scientific literacy of college graduates has
remained stagnant for many years (Johnson and Pigliucci 2004; Impey et al. 2011; Walker et al.
2002). Traditional introductory chemistry courses have been shown to fail to raise student critical
thinking skills to necessary levels. It has also been shown that student collaboration improves
critical thinking skills (Gokhale and Anuradha 1995).

For this project, the author used Bloom’s Taxonomy of the Cognitive Domain (Bloom 1956) to evaluate a series of in-class case studies given to students in a second-semester General Chemistry course. Bloom’s Taxonomy is shown in Figure 1 below.

![Bloom's Taxonomy Scheme](image)

**Figure 1: Bloom’s Taxonomy Scheme (Bloom 1956)**

Bloom’s Taxonomy is a pedagogical tool consisting of six categories (i.e., Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation) corresponding to increasingly complex orders of thinking (Bloom 1956). Knowledge corresponds to learning specific facts and information, while Comprehension might include taking the information and paraphrasing the explanation into a new form. Application involves solving problems using established methods or equations, while Analysis gleans meaning from the dataset. The synthesis category involves a novel approach to analyzing a problem and drawing conclusions, while Evaluation involves the
critical analysis of the approach and its contribution to knowledge (Bloom 1956). Bloom’s Taxonomy is a useful tool for calibrating the complexity of classroom lessons so they correspond with the expected experience and knowledge bases of the students. The author hypothesized that students in the course facilitated by case studies would exhibit a higher level of thinking and thus score higher on the higher tier classified questions than the students in courses without case studies.

**Methods**

In-class case studies were used in a Majors second semester general Chemistry course to encourage students to use and develop these critical thinking skills. The first case study, entitled ‘Can of Bull,’ had students identify ingredients in different energy drinks and use their knowledge of biomolecules to determine whether the drinks provided actual energy or merely stimulation. The second case study, entitled ‘Iodine Clock,’ had students use a cyclic ‘clock reaction’ to determine how reaction kinetics depend on different variables. In the third case study, entitled ‘One Headache after Another,’ students used real life examples of buffers in the human body to solve a medical problem. Students used the patient’s blood test and their critical thinking skills to discover the buffer system found in the blood and how it is affected by change such as a decrease of carbon dioxide concentration. In the fourth case study entitled, ‘Rubber Band’, students determine reaction spontaneity using enthalpy, entropy and Gibb’s free energy. In the fifth case study, entitled ‘Real World REDOX,’ students attempt to solve the problem of acid mine drainage using their knowledge of REDOX chemistry and solubility. In the final case study, entitled ‘Ni(II) complex ions,’ students relate observations of a solution’s color, caused by d-orbital energy splitting, to ligand exchange.
Student performance was monitored by a pre/post concept survey with questions requiring various levels of thinking. This survey was originally fabricated to measure learning gains for a separate study. It was useful for this study for two reasons: it provides historical data for comparison and it has a variety of questions which can be classified to measure student critical thinking skills. The pre/post concept survey was also given to the Non-Majors second semester Chemistry course. This course taught the same material in the same period of time as the Majors course. Thus it was used as a control group for this experiment.

There were 48 students enrolled in the study for 5 months. In-class case studies were implemented biweekly by students working in groups. During this time a camcorder was set up in the front of the classroom to obtain an overview of class structure. Observations during in-class case studies were recorded using an observation worksheet with a classroom map to monitor students’ use of critical thinking throughout the activity. Student results on critical thinking intensive questions on 3 midterm exams and the final exam monitored student ability throughout the semester.

In order to quantify students’ critical thinking skills, a Bloom’s taxonomy classification scheme was used. The position on the pyramid depicted in Figure 1 corresponds to the required cognitive complexity where higher tiers correspond to higher levels of thinking. This scheme was used to classify concept survey questions as well as student discourse during classroom observation. Each of the questions on the concept survey was grouped into one of the first four levels of Bloom’s pyramid since these are the levels of thinking applicable to an introductory college chemistry course. This classification was performed by Chemistry faculty members and graduate students to decrease classification bias.
Results of Research Project

To quantify the Pre/Post Concept Survey data, the percent of students in each course who answered a question correctly was calculated (Eqn. 1).

$$\% \text{ Correct} = \frac{\# \text{ students with correct answer}}{\text{total } \# \text{ students}} \times 100$$

Equation 1

Using the value calculated in Equation 1, the average percent of the percent of students in each course who answered a question correctly for a particular level on Bloom’s pyramid could be calculated (Eqn. 2).

$$\text{Level Avg } \% = \frac{\text{Sum(\% Correct for question for this level)}}{\text{# questions in this level}}$$

Equation 2

Using the value calculated in Equation 2, the difference between the pre and post concept survey average percent for each level could be calculated (Equation 3).

$$\text{Level } \% \text{ Difference} = \text{Level Avg } \% \text{ Post} - \text{Level Avg } \% \text{ Pre}$$

Equation 3

These differences are listed in Table 1 below and can be used as a quantitative comparison between the controls and experiment.
The author performed a t-test ($\alpha=0.95$) on the data obtained using Equation 2 to test the null hypothesis (i.e. that the mean percent correct for each Bloom’s level in Majors 2015 compared with Majors 2014 and Non-Majors 2015 have no significant difference.) The only significant results for the pre/post concept survey were for the Knowledge and Application levels of Bloom’s Taxonomy between the Spring 2015 Majors course with case studies and the Spring 2015 Non-Majors course without case studies ($p=0.006$, $p=0.04$). However, the spring 2014 Majors course without case studies also demonstrates a significant difference in these levels with the spring 2015 Non-Majors course. This suggests that the difference is not due to the implementation of in-class case studies.

It should be noted that the concept survey was designed to assess learning gains, not critical thinking skills. Since this survey is historically given to all second-semester General
Chemistry courses, it provided a historical comparison for the current study. A survey that is well-designed and validated for assessing critical thinking skills given over multiple semesters may provide a more accurate assessment.

There was no significant difference between the short answer portions of the midterm exams given in the spring 2015 and 2014 Majors courses. However, the material was taught in a different order in spring 2015 than in spring 2014. Thus the only relevant comparison data is the short answer portion of the final exam for each year. The difference between the spring 2015 and spring 2014 short answer final exam scores was not statistically significant.

The qualitative in-class observation data were also inconclusive. Students’ levels of thinking classified by Bloom’s taxonomy varied depending on the case study at hand but did not illustrate a specific trend. This could be due to the nature of the case studies themselves. The questions in the case study worksheets completed by students were classified using Bloom’s taxonomy to look for a correlation between the Bloom’s classification of the worksheet and observed discussion by students. The percentage of each level present in both observation and worksheet classification for each case study is given in Table 2 below:
Table 2: Each horizontal colored bar represents a different case study in chronological order. In-class observation describes the percentage of student statements which were classified by a particular Bloom’s level within a group of students. Worksheet classification indicates the percentage of questions on each student worksheet which were classified by a particular Bloom’s level. Classification was done solely by the author.

There does not appear to be a consistent relationship between student discussion and worksheet question classification. This result suggests that the worksheet content was not the determining factor in student thinking. However, the in-class observation data were obtained by
a single observer and only reflects a single group of students whose composition varied slightly throughout the semester. Multiple observers and in-depth analysis of student discussion from isolated audio recordings of student groups would give a more accurate depiction of in-class student discussions.

Discussion

In conclusion, this study indicates that students in the Majors course exhibit a greater aptitude for higher level critical thinking skills than students in the Non-Majors course. This is demonstrated by Majors students’ higher proficiency in Application questions compared to Non-Majors’ students. However, a similar proficiency is demonstrated by the Majors course without the implementation of case studies which suggests that the higher proficiency is not due to case study implementation. Nevertheless, there is a degree of uncertainty in these results since critical thinking skills usually develop on a longer timescale than the duration of this study. Although students’ critical thinking skills do not appear to improve after implementation of in-class case studies based on the methods used in this study, a more thorough study over a longer period of time is necessary to truly probe the effect on students’ critical thinking skills. Future studies will include larger sample sizes with multiple in-class observers to analyze student interactions. We also recommend examining the course content using Bloom's to improve the fit between worksheets and classroom interactions.
References Cited


Chapter Six
Quantifying the Impact of a Physics Co-Seminar
Daniel C. Cole
Department of Physics

For my Teaching-as-Research project, I chose to investigate my own teaching. The physics co-
seminar course that I teach gives me an opportunity that is somewhat unique for a graduate
student in the physical sciences – I develop and deliver my own curriculum. A TAR project
presented a good chance to look critically at what I do and how it could be improved. I chose to
apply some assessment tools developed by the field of Physics Education Research to investigate
the impact my course has on my students.

I am the lead instructor for a physics co-seminar course run by the University of Colorado’s
Student Academic Success Center (SASC). The Student Academic Success Center supports low-
income and first-generation students at CU from a variety of diverse backgrounds by providing a
supportive, multi-cultural learning community. SASC represents a key component of CU’s
efforts to meet an important national goal laid out by a 2012 report from the President’s Council
of Advisors on Science and Technology (PCAST): to improve undergraduate science outcomes,
particularly among students from communities historically underrepresented in higher education,
by improving the first two years of education in science, technology, engineering, and
mathematics (STEM) fields, providing students with the tools to excel, and diversifying the
pathways to STEM degrees (PCAST 2012). SASC’s support for students in STEM courses
aligns with the PCAST report’s call for a thirty-four percent increase in the annual number of
undergraduate STEM degrees awarded.
SASC supports STEM students from its target population in two primary ways: when possible, SASC runs its own introductory STEM courses with smaller class sizes and more frequent class meetings. When this is not possible, SASC supplements large-lecture introductory STEM courses with a co-seminar course. The co-seminars provide students taking introductory STEM classes with extra academic support. A primary goal of the co-seminar program is to increase the retention of low-income and first-generation students in introductory STEM course sequences, which often serve as a barrier to entry for their chosen field. SASC’s co-seminars function as small, supportive learning environments, and focus on using research-based pedagogical strategies to help students meet their academic goals. Students enroll concurrently in a co-seminar run by SASC and the corresponding large-lecture introductory STEM class run by the STEM department. As the lead instructor for the physics co-seminar, I independently design curriculum to help my students develop their understanding of physics concepts – my most important, immediate goal is to help students gain the knowledge they need to succeed on the homework assignments and exams from their larger introductory physics course so that they can meet their academic goals.

SASC’s own internal research shows that students enrolled in co-seminars are more likely to meet their academic goals than students from the same demographics who are not enrolled in the co-seminars – this is reflected in final course grades and in retention within a student’s desired major. In the specific case of the physics co-seminar there is an opportunity to use additional tools developed in the context of Physics Education Research to evaluate the effectiveness of the program.

The field of Physics Education Research takes a quantitative approach to developing techniques for measuring and improving the effectiveness of physics teaching. The field has
blossomed in the past three decades, with one major focus on developing standardized tools to measure how physics instruction impacts students’ problem-solving abilities, conceptual understanding, and beliefs about science. These tools make these measurements in a way that is independent from the metrics which evaluate students in their courses such as exam scores, performance on homework assignments, laboratory projects, and so on. The field has developed and validated several *conceptual assessments* – tools for measuring understanding of physics concepts – such as the Force-Motion Conceptual Evaluation (FMCE) (Thornton and Sokoloff 1998), the Force Concept Inventory (Hestenes et al. 1992), and the Brief Electricity and Magnetism Assessment (Ding et al. 2006). The field has also developed and validated *attitudinal assessments* – tools for measuring students’ attitudes and beliefs about the nature of science, scientific inquiry, and the practice of physics. The attitudinal assessments that are used most frequently are the Maryland Physics Expectations Survey (MPEX) (Redish et al. 1998) and the Colorado Learning Attitudes about Science Survey (CLASS) (Adams et al. 2006).

In my Teaching-As-Research project, I used the FMCE and the CLASS to compare the learning gains (as measured by the FMCE) and evolution towards or away from expert-like beliefs about physics (as measured by the CLASS) of three groups of students: those enrolled in my co-seminar, students participating in other SASC programs and taking introductory physics but not enrolled in my co-seminar, and students not affiliated with SASC taking the same introductory physics courses.

In this article, I will describe my TAR project in detail, including the background necessary to understand the results. For context, I will begin with a description of my co-seminar pedagogy. I will then discuss the background for the FMCE and CLASS assessments. I’ll discuss the methods I used to carry out my TAR project, and the results of the project. In particular, the
students in my co-seminar showed shifts on the CLASS which were significantly better than those shown by the other two student cohorts. I will connect this result to specific components of my co-seminar curriculum. Finally, I will discuss possible directions for continued research.

The SASC Physics Co-seminar

My primary goal in teaching the physics co-seminar is to help students build their understanding of physics so that they can get the grade that they want in their introductory physics class. The co-seminar curriculum I use has four key components: in-class student-led group problem solving, ‘whiteboard time,’ an in-class quiz that precedes each exam in the large-lecture physics course, and two homework assignments due weekly: the Lecture Reflection and the Homework Analysis. For the semester in which I conducted my TAR project, I was supported in the classroom by an Instructional Assistant (IA), a senior-level undergraduate engineering student. I had 11 students in my class, all concurrently enrolled in the CU Physics Department’s calculus-based introductory physics course.

The bulk of in-class co-seminar time is spent on group problem solving. Problems are taken from the student workbook accompanying Randall Knight’s *Physics for Scientists and Engineers: A Strategic Approach* (Knight 2004), a PER-based introductory physics text, and from the University of Minnesota’s online archive of Context-Rich Problems (“Context-Rich Problems” 2017), which challenge students to identify and apply the correct physical principles in the context of a real-world scenario. I also develop worksheets to treat specific concepts as the need arises. My IA and I focus on facilitating student discussion during group problem-solving time, so that the students learn from and teach each other.
The other significant component of class time is what we refer to in class as “whiteboard time” which is often improvisational and responds directly to students’ needs. Students ask me to address conceptual questions that arose for them either during their lecture course, while doing homework, or while preparing for an exam, and sometimes I will deliver a short impromptu lecture. This is effective because I respond in real-time to questions and concerns, which is an important contrast with the situation in their large-lecture course. Other times I call a student volunteer to the whiteboard, and with my guidance and this student’s leadership, the class will work together as a team to investigate a particularly troublesome concept or solve a tricky problem.

We spend a relatively small fraction of in-class co-seminar time on a quiz which precedes each exam in the large-lecture physics course. This quiz serves as an important formative assessment for the students, and it also gives them practice solving physics problems on their own in a setting that more closely mimics exam conditions than studying on their own or with friends. My students have indicated that these quizzes are helpful because they reveal gaps in their knowledge in advance of the exam, and also because they provide an opportunity to practice test-taking strategies and prepare to deal with exam-day nerves.

Outside of the classroom, students complete two homework assignments each week: a Lecture Reflection and a Homework Analysis. The goal of these homework assignments is not as much to develop conceptual understanding, a goal which is already served by the homework assigned in their physics course, as to give students a chance to practice and develop metacognition skills. Metacognition is an important concept in PER (and in education theory in general), and refers to everything that goes on in solving a problem or learning a subject besides the actual assimilation and application of new conceptual material. It can consist of asking
oneself “Does this make sense to me?” or “How should I start this problem?” or of making a plan, evaluating whether one’s answer is reasonable, or critically assessing one’s note-taking strategy to determine whether it generates useful course notes. An important work on metacognition which includes more background is the essay “What’s All the Fuss About Metacognition” by Alan Schoenfeld (1987). Broadly, metacognition can be described as having three facets: self-awareness, self-regulation, and the beliefs and intuitions which form the context for new information (Schoenfeld 1987). I emphasize metacognition here because it is closely related to the attitudes and beliefs assessed by the CLASS.

The Lecture Reflection is roughly one typed-page long and consists of, first, a summary of the lecture content for the week, and then a reflection portion in which students write critically about how class is going for them, whether their study practices are benefitting them, what concepts they’ve found challenging, and any other course-related observations or concerns that they have. The Lecture Reflection requires students to turn off ‘auto-pilot’ with respect to their physics course and think about how to use their time efficiently and effectively. It also gives me more insight into what concepts students need additional help with.

For the Homework Analysis, students choose one homework problem that they completed for their large-lecture physics course during the past week to explore in more detail. After summarizing the problem and their solution to it, they are asked to complete a series of higher-level thinking tasks: 1. Evaluate whether their answer is reasonable, 2. Identify the key physical concepts used in their solution to the problem, 3. Tell the story of the problem in prose – what happens and why, 4. Locate their problem solving strategy on Bloom’s taxonomy (Bloom et al. 1956), and 5. Using the same concepts, write a new problem and solve it. The Homework Analysis is intended to counter students’ tendency to rush through their homework to reach the
answer as quickly as possible and, once the answer is reached, to move on without more thought. Item number 3 is especially important to my pedagogical priorities because it requires students to use words to describe physical concepts, which requires a level of understanding that is above application of equations from the textbook. Further, it encourages students to think about cause and effect, a relationship which a solely equation-focused problem-solving style can obscure.

The FMCE Conceptual Assessment

The Force and Motion Conceptual Evaluation is a ‘research-based, multiple-choice assessment of student conceptual understanding of Newton’s Laws of Motion’ developed by Ronald Thornton and David Sokoloff and published in 1998 (Thornton and Sokoloff 1998). The FMCE followed the development of the Force Concept Inventory (FCI) in Hestenes et al. (1992). To put these assessments and the field of PER in general into context, it is instructive to directly quote from the beginning of the article introducing the FCI:

Specifically, it has been established that (1) commonsense beliefs about motion and force are incompatible with Newtonian concepts in most respects, (2) conventional physics instruction produces little change in these beliefs, and (3) this result is independent of the instructor and the mode of instruction. The implications could not be more serious. Since the students have evidently not learned the most basic Newtonian concepts, they must have failed to comprehend most of the material in the course. They have been forced to cope with the subject by rote memorization of isolated fragments and by carrying out meaningless tasks. No wonder so many are repelled! The few who are successful have become so by their own devices, the course and the teacher having supplied only the opportunity and perhaps inspiration (Hestenes et al. 1992:141).

Much of the field of PER has focused on trying to solve the problem presented here, and the FCI and FMCE provide a universal way of measuring progress towards that goal. The FMCE consists of 47 multiple-choice questions testing concepts from introductory mechanics. Incorrect answer
choices deliberately match common student misconceptions, although as the authors of the FCI point out, calling these ideas ‘misconceptions’ is a bit unfair, because they are ‘significant commonsense beliefs [that] have been firmly held by some of the greatest intellectuals in the past… They happen to be false, but that is not always so easy to prove” (Hestenes et al. 1992: 141). Overcoming and replacing these misconceptions is a key challenge in effective physics instruction, and the FMCE is an important metric to use to evaluate one’s success.

The standard procedure for using the FMCE is to administer it pre-instruction, at the beginning of a semester, and post-instruction. Once this is done, Hake’s normalized gain (Hake 1998) is a useful statistic that facilitates comparison across classes and student populations by making an attempt to control for the knowledge of the incoming class. Hake’s gain is defined as:

\[ g = \frac{\langle \text{post} \rangle - \langle \text{pre} \rangle}{100 - \langle \text{pre} \rangle} \]

where \( \langle \text{post} \rangle \) and \( \langle \text{pre} \rangle \) denote class-averaged post- and pre-test scores in percent. Hake’s gain, in words, is the amount that was learned divided by the amount that could have been learned. It’s worth emphasizing that this quantity is a gain of averages, not an average of gains.

A key result in each of the seminal papers by Thornton and Sokoloff (1998), Hestenes, Wells, and Swackhamer (1992), and Hake (1998) is that most of the time, traditional physics instruction does not help students learn physics if they would not have been able to learn it on their own. Whether or not a new pedagogical technique performs better than the baseline represented by traditional physics instruction, as indicated by the FCI or the FMCE or some other material-appropriate conceptual assessment, has become a key question in PER in evaluating new pedagogy. It was my goal to apply the FMCE to investigate the effect of taking
both my co-seminar and the introductory physics course at CU, relative to taking only the introductory physics course.

The CLASS Attitudinal Assessment

The purpose of an attitudinal assessment is to measure to what extent students’ beliefs and expectations about the practice of science, scientific problem-solving, and the nature of scientific knowledge agree with the beliefs and expectations of practicing scientists, referred to as ‘expert-like beliefs.’ As Alan Schoenfeld notes, “People are *interpreters* of the world around them… [they] perceive what they experience in the light of interpretive frameworks they have developed” (Schoenfeld 1987:195). For this reason, it is intuitive that attitudes about science and the development of conceptual understanding are tightly linked in an introductory physics class. But there are other reasons that beliefs about science are important, and a particularly significant one is that for many students the development of scientific literacy is a more important success outcome of an introductory physics course than skill in physics problem-solving. Students may never again need to solve a physics problem after taking introductory physics, but it is very likely that in their professional or private lives they will receive science-based recommendations and counsel, and having an understanding of the nature of scientific knowledge and practice is important.

If we seek to foster expert-like attitudes about science in our physics students, and to determine whether we are doing so, then it is natural to use a longitudinal measurement scheme with pre- and post-instruction tests. Borrowing a general term from the sociology of education, the evolution of a student’s beliefs about science and physics problem-solving over the course of a semester in an introductory physics class can be attributed to the “hidden curriculum,” the set
of implicit lessons taught by a course about the discipline. These lessons come not from explicit instruction but from, for example, a mismatch between the expectations of the instructor and the expectations of the student, or from a difference in objectives. Edward F. Redish points out in his textbook on teaching physics that “Most of our students don’t know what you and I mean by ‘doing’ science or what we expect them to do” (Redish 2003:52), and he explains that this misunderstanding can “inadvertently wind up encouraging students to hold unfavorable attitudes” (Redish 2003:55).

Use of attitudinal assessments, first the MPEX and then the CLASS, has revealed that students’ beliefs often shift away from expert-like beliefs during introductory physics instruction, even in courses where students demonstrate strong learning gains as measured by e.g. the FMCE or FCI (Madsen et al. 2015). In short, instructors need to pay attention to their hidden curriculum – either by being aware of and managing the lessons that their course requirements are teaching to students about how to study and learn physics, or by actually implementing explicit curriculum designed to foster expert-like beliefs.

The CLASS consists of 42 prompts with responses on a 5-point Likert scale ranging from strongly agree to strongly disagree, with a neutral middle response. Some prompts are:

- A significant problem in learning physics is being able to memorize all the information I need to know.
- Knowledge in physics consists of many disconnected topics.
- To understand physics I discuss it with friends and other students.

The CLASS differs from previously developed conceptual assessments such as the MPEX in several ways, with the most significant being that “The statements were written to be meaningful even to students who had never taken physics,” “The ‘expert’ and ‘novice’ responses to each statement were unambiguous so scoring of the responses was simple and obvious” and finally
“The amount of time required to thoughtfully complete the survey was kept to ten minutes or less” (Adams et al. 2006). This last point has led to the CLASS’s widespread adoption across many physics programs, and helped in my efforts to recruit a sufficiently large sample size for this project. Another important aspect of the CLASS’s development is that the questions have been empirically placed into categories based on measured correlations between responses to different questions, which is “in contrast to a priori groupings of statements by the survey creators based on their belief as to which statements characterize particular aspects of student thinking” (Adams et al. 2006:1). The empirically-determined categories, and an example statement for each, are:

1. Real world connections
   - Learning physics changes my ideas about how the world works.
2. Personal Interest
   - I am not satisfied until I understand why something works the way it does.
3. Sense making/effort
   - There are times I solve a physics problem more than one way to help my understanding.
4. Conceptual understanding
   - Knowledge in physics consists of many disconnected topics.
5. Applied conceptual understanding
   - A significant problem in learning physics is being able to memorize all the information I need to know.
6. Problem solving general
   - I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.
7. Problem solving confidence
   - If I get stuck on a physics problem, there is no chance I’ll figure it out on my own.
8. Problem solving sophistication
   - If I don’t remember a particular equation needed to solve a problem on an exam, there’s nothing much I can do (legally!) to come up with it.
Not all 42 questions are categorized under one of these 8 headings. This happens when responses to a particular question don’t correlate strongly with responses to a cluster of questions. Finally, some questions fall into multiple categories.

**Data Collection for the Teaching-as-Research Project**

For my TAR project, I collected pre- and post-instruction FMCE and CLASS assessments from three cohorts of students: SASC students in my co-seminar, SASC students (i.e. students participating in any SASC program) not in my co-seminar but taking introductory physics, and non-SASC students taking introductory physics. I administered the CLASS online using Google Forms, and analyzed the data using the scoring sheet provided on the University of Colorado’s website about the CLASS (CLASS 2017). I solicited participation from my students through in-class announcements and from the other two cohorts by email with the help of my SASC supervisor and research partner, Dr. Becca Ciancanelli, and the instructor for the large-lecture physics course, Dr. Mike Dubson. I obtained informed-consent at the time of response to the survey with a form that preceded the CLASS. Students in both SASC cohorts were offered a $5 gift card for participation in the study, which was funded by SASC. Students in my co-seminar were also offered extra credit for participating in the study or indicating in the online portal for the study that they did not wish to participate. This alternative is important to eliminate the possibility that my students could have felt coerced to participate in the study. To further protect the students, the identities of the students who participated were kept from me until after I submitted final grades for the semester – Dr. Ciancanelli assisted in this by receiving the online responses and delivering them to me after the semester was over. Students unaffiliated with SASC in the general introductory physics cohort were not compensated for taking the CLASS.
Students provided a unique identifier when they took the CLASS online, which allowed me to match pre- and post-instruction data and calculate shifts for each student.

The CU physics department administers the FMCE at the beginning and end of each semester in the introductory physics recitation sections, so I did not administer it myself. I obtained consent from both SASC cohorts for the physics department to share their FMCE scores with me through the informed-consent form preceding the CLASS survey. For the third, non-SASC cohort, the physics department simply shared class-average FMCE statistics with me, which did not require any special permissions or consent.

I described these procedures in detail in a document submitted to CU’s Internal Review Board (IRB), and my project was approved before I began soliciting study participation from students. The IRB process went smoothly for me because I thought quite carefully with Dr. Ciancanelli about how to minimize risks to students and maximize the chances for a successful project before submitting the proposal. Writing the IRB document was helpful in itself, because it required me to put all of my plans into a formal document which I could then follow as I carried out the project.

**FMCE Results**

I will first briefly discuss the results of the FMCE portion of my project, then move to a longer discussion of the more substantive CLASS results. For the FMCE data, the size of my SASC co-seminar (non-co-seminar) cohort was N=6 (N=4). However, following FMCE analysis practice that is standard for the CU physics department of removing pre- and post-tests where too many questions are left blank reduced the sizes of the FMCE cohorts to N=4 and N=2.
Using Hake’s learning gain \( g \) defined above, the normalized gain of averages for the co-seminar cohort was 8.3 %. The gain of averages for the non-co-seminar SASC cohort was 51 %, while the gain of averages for the entire introductory physics course was 47 %. In calculating the gain of averages for the co-seminar cohort, I removed the data for a student who had an extremely high pre-test score (91%) and scored slightly lower on the post-test, because this data point is an obvious outlier.

It is troubling that the learning gain for the co-seminar cohort is so low, both in absolute terms and relative to the other two cohorts. Taken at face value, this result suggests that the co-seminar class impedes development of conceptual understanding. However, the final sample sizes were \( N=3 \) and \( N=2 \) for the SASC co-seminar and non-co-seminar cohorts respectively, after removing the outlying data point from the co-seminar cohort. The strength of any conclusion is obviously very weak, because with these sample sizes it is unlikely that the samples are representative of the respective populations. Further, there is no obvious mechanism by which the co-seminar would hamper student learning, and this conclusion is contradicted by feedback from students and by SASC’s other internal studies of the effectiveness of the co-seminar. It would be unjustified to write off these results entirely, and they suggest that further investigation is warranted, perhaps with a way to recruit more students to participate in the study. On the other hand, these results don’t form a strong basis for the conclusion that the co-seminar is ineffective.

**CLASS Results**

The CLASS data I collected is presented in Table 1. For the CLASS data, the sizes of my co-seminar, non-co-seminar SASC, and non-SASC cohorts were 7, 4, and 22 respectively. Students
volunteered to participate in the study, so there could be selection bias effects at work. However, this does not appear to be relevant to the particular conclusions drawn from the study. I included in the data set a student who completed three-quarters of the semester in the co-seminar and the introductory physics course before withdrawing from both. This student responded to the post-instruction CLASS survey at the end of the semester.

In Table 1, the ‘Overall’ row shows the data for all questions, while the ‘All categories’ row excludes those questions which don’t fall into an empirically-defined category. For each question, a response is favorable (‘F’ in the table) if it is the expert-like response, or unfavorable (‘U’ in the table) if it disagrees with the experts’ response. Responses can also be neutral. In analyzing CLASS responses, ‘Agree’ and ‘Strongly agree’ responses are treated identically, and ‘Disagree’ and ‘Strongly disagree’ responses are similarly grouped together. For each category, the table shows the shifts from pre-test to post-test for each student in the percentage of total in-category responses which were favorable or unfavorable, averaged over the students who completed every question for the category. The favorable and unfavorable shifts do not sum to zero because shifts may be to or from the neutral response. The table also shows the net shift, which is calculated as the difference between the favorable and unfavorable shifts. To assist with reading of the table, shifts towards expert-like beliefs are colored green, and shifts away from expert-like beliefs are colored red.

We can immediately note that the SASC co-seminar cohort had (small) favorable net shifts overall and in the Personal interest, Real world connection, Sense-making/effort, and Conceptual understanding categories. Further, in categories having net unfavorable shifts, the shifts are smaller in magnitude than either of the other two cohorts for every category except Problem solving confidence.
Table 1. Summarized results of the CLASS conceptual assessment across the three student cohorts considered in this research. Shifts in the direction of expert-like beliefs are colored green, and shifts away from expert-like beliefs are colored red. Bold type indicates statistical significance at the 95% confidence level.

<table>
<thead>
<tr>
<th></th>
<th>SASC physics co-seminar</th>
<th>SASC no co-seminar</th>
<th>Non-SASC, general population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=7</td>
<td>N=4</td>
<td>N=22</td>
</tr>
<tr>
<td>F U Net (F-U)</td>
<td>F U Net (F-U)</td>
<td>F U Net (F-U)</td>
<td>F U Net (F-U)</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.6 -1.9 1.3</td>
<td>-8.3 -0.7 -7.6</td>
<td>-11.7 6.5 -18.2</td>
</tr>
<tr>
<td>All categories</td>
<td>1.1 -4.4 5.5</td>
<td>-14.4 4.8 -19.2</td>
<td>-14.1 7.1 -21.2</td>
</tr>
<tr>
<td>Personal Interest</td>
<td>7.1 -7.1 14.2</td>
<td>-12.5 0 -12.5</td>
<td>-15.9 10.6 -26.5</td>
</tr>
<tr>
<td>Real world connection</td>
<td>3.6 -3.6 7.2</td>
<td>-12.5 0 -12.5</td>
<td>-11.4 3.4 -14.8</td>
</tr>
<tr>
<td>Problem solving general</td>
<td>-7.1 -1.8 -5.3</td>
<td>-15.6 6.3 -21.9</td>
<td>-17 14.8 -31.8</td>
</tr>
<tr>
<td>Problem solving confidence</td>
<td>-17.9 3.6 -21.5</td>
<td>-12.5 6.3 -18.8</td>
<td>-23.9 20.5 -44.4</td>
</tr>
<tr>
<td>Problem solving sophistication</td>
<td>-14.3 2.4 -16.7</td>
<td>-16.7 12.5 -29.2</td>
<td>-22 15.2 -37.2</td>
</tr>
<tr>
<td>Sense-making/effort</td>
<td>6.1 -10.2 16.3</td>
<td>-7.1 -7.1 0</td>
<td>-14.9 4.5 -19.4</td>
</tr>
<tr>
<td>Conceptual understanding</td>
<td>9.5 -2.4 11.9</td>
<td>-16.7 16.7 -33.4</td>
<td>-13.6 7.6 -21.2</td>
</tr>
<tr>
<td>Applied conceptual understanding</td>
<td>-2 2 -4</td>
<td>-17.9 21.4 -39.3</td>
<td>-15.3 4.2 -19.5</td>
</tr>
</tbody>
</table>

Before making an attempt to interpret my CLASS data, it’s worth considering some statistics. Note that my sample sizes are very small, with 7 respondents in the co-seminar cohort. As a result, only some of the shifts I observe are statistically significant, which in this case means that, under the assumption of individual shifts which are normally distributed about the mean for the population, there is at least 95% chance that the mean for the population is different from zero. Statistically significant shifts are indicated in the table by bold type. For a given category, statistical significance of the shift is determined as follows, using percent favorable responses as an example: the shift within the category is determined for each student as

$$\delta_i = 100 \times \frac{N_{Favorable,post} - N_{Favorable,pre}}{N_{category}}$$

where $i$ indexes over the students and $N_{Favorable,post}$, for example, denotes the number of favorable responses to questions in the category given on the post-test. For the category, the
mean $\bar{\delta}$ and standard deviation $\sigma$ of the shifts $\delta_i$ are calculated, and the uncertainty $\sigma_{\bar{\delta}}$ in the mean for the population is the standard deviation divided by the square root of the sample size:

$$\sigma_{\bar{\delta}} = \sigma / \sqrt{N}$$

An observed shift is statistically significant if it is larger than twice this uncertainty.

There are two ways to consider the data revealed by the CLASS component of my research: The first way is to consider only the overall shift in percent favorable responses, -0.6, within the context of the studies considered in a meta-analysis by Madsen et al. (2015). The meta-analysis compiled results from 24 different studies, and considered the overall shift in percent favorable responses for each of these studies. The authors performed analysis-of-variance (ANOVA) tests to determine the correlation between each of teaching method, class size, and student population with shifts in beliefs measured by the CLASS/MPEX. The ANOVA revealed that each of these three attributes had an effect which was statistically significant at or better than the 99% confidence level, with teaching method having the largest – the F-statistics for teaching method, student population, and class size were 27.4, 15.8, and 5.4 respectively. When grouping data points according to a given attribute, the F-statistic is the ratio of the variance between groups to the variance within groups. A higher F-statistic indicates that grouping according to a given attribute more effectively explains the data.

The mean overall shifts in percent favorable responses for students in the two non-co-seminar cohorts are typical of students in general introductory physics courses with a teaching method where no special attention is paid to developing expert-like beliefs about science and the practice of physics. The results of the students in the SASC physics co-seminar are typical of students in courses ‘with “some” focus on developing expert-like beliefs,’ which, as reported in the meta-analysis, have been observed to have mean shifts to percent favorable responses
between -1 and 2. This is consistent with my teaching method – my main goal was not to explicitly focus on fostering expert-like beliefs about physics, but to help students build the knowledge they needed to succeed on their homework assignments and exams in their introductory physics course. However, I did place some emphasis on developing metacognitive skills with the Lecture Reflection and Homework Analysis assignments, encouraging students to think about how they learn physics and solve physics problems. Further, I suspect that ‘whiteboard time’ contributed significantly to students’ development of more expert-like beliefs about problem-solving in physics. For example, sometimes a student would ask me a question I did not know how to answer immediately, and I think that explicitly modelling for the students the process of working out a solution in the face of this uncertainty was beneficial for their understanding of how problem-solving in physics works.

The experience of my co-seminar students differs from the students considered in the studies included in the meta-analysis in an important way. If we consider teaching method to be a treatment variable, then they are a group that was given multiple treatments, as they were enrolled concurrently in their large-lecture introductory physics course and my co-seminar. The ‘some attention to developing expert-like beliefs’ teaching method used in my co-seminar appears to have been more important in determining their overall CLASS shifts than the ‘no attention to developing expert-like beliefs’ teaching method used in the large-lecture introductory courses, because their CLASS shifts are clearly comparable to students in the former group. This has practical implications – it suggests that in order to foster expert-like beliefs, or at least avoid pushing students in the other direction, it may be sufficient to provide just a little supplemental instruction instead of reforming a course’s entire curriculum.
A second way to consider the CLASS data I collected is to consider the breakdown of the shifts into the 8 empirically-determined categories. It is encouraging that net shifts were favorable in the Personal interest (14.2), Real world connection (7.2), Sense-making/effort (16.3), and Conceptual understanding (11.9) categories. The net shifts for co-seminar students were unfavorable in the three categories related to problem-solving: Problem solving general (-5.3), Problem solving confidence (-21.5), and Problem solving sophistication (-16.7). The net shifts were also unfavorable in the Applied conceptual understanding category (-4). One way to interpret this data is with the hypothesis that at the time of the post-test my students were beginning to develop expert-like beliefs, but that the process was not complete. They were interested in learning physics, and looked for examples of what they were learning in the world around them. They understood that learning physics would take hard work, and requires developing an understanding of concepts, in contrast with the simple memorization of equations. It seems natural to me that development of these beliefs would precede the development of an expert-like approach to problem solving, which requires a sophisticated application of the ideas above to solving problems which can be, at times, frustrating and confusing. While the sample sizes in my study are too small to draw firm conclusions, the data suggests that it is worth exploring in future research whether this particular sequence of events is universal, or at least a recurring path to an expert-like attitude: that first students come to understand physics as a discipline in broad strokes, and then develop the ability to apply this understanding to the nitty-gritty process of solving problems. The data also suggests that I should place more attention on developing good problem-solving attitudes and habits in my pedagogy, perhaps by explicitly and deliberately modelling the process for students more frequently than I do, and by creating
structured worksheets designed to take students step-by-step through a problem the way an expert would solve it.

**My TAR Experience and Future Directions**

Conducting this Teaching-as-Research project taught me about performing social science research, and gave me insight into my own teaching. The project presented an opportunity to put into practice knowledge I gained from taking CU’s ‘Teaching and Learning Physics’ course, and from participating in the 2016 Research on Academic Retention symposium put on by CU’s Teaching Institute for Graduate Education Research. I plan to pursue a career teaching college-level physics, and having at my disposal the tools I learned about through this TAR project will be invaluable in evaluating my own teaching, and in enabling me to make use of results from and contribute to the field of Physics Education Research. In particular, going through the process of submitting a proposal to the Internal Review Board was eye-opening, since this is not required for my primary research in laser science. Social science makes use of statistics differently from physical science, and so applying statistics to understand my data was also educational. The data I collected in the FMCE portion of my study is important, although it is hard to draw firm conclusions from it. This data may indicate that my teaching of the co-seminar is ineffective, and that the co-seminar is not helping students develop conceptual understanding. I strongly suspect that this is not the case, based on my experiences teaching and on feedback I’ve received from students. The data certainly indicates that further study is warranted, and underscores for me the importance of having a sufficiently large, representative sample.

This TAR project demonstrated that my teaching has an effect on students’ beliefs about physics that is better than the large-lecture introductory course without the co-seminar. This is
encouraging, because it tells me I’m doing something right! My CLASS results also highlight the particular areas where my pedagogy could be improved. Being able to use widely-accepted tools to come to this conclusion is exciting. In addition to the obvious implications that further study of FMCE conceptual gains with a representative sample is warranted, there are two other exciting directions for further research suggested by my CLASS results: 1. An investigation into the development of supplemental pedagogy which helps to foster expert-like beliefs without requiring a full-blown overhaul of primary teaching method, and 2. An investigation into the process of the development of expert-like beliefs among physics students with finer temporal resolution, in contrast with a simple pre-test and post-test.

I am grateful to my SASC supervisor Dr. Becca Ciancanelli for partnering with me in this research, to Professors Noah Finkelstein and Steve Pollock for helpful discussions and guidance, and to Dr. Mike Dubson for assistance soliciting CLASS volunteers. Their support, along with the support of CU Boulder’s Graduate Teacher Program and the Teaching institute for Graduate Education Research (TIGER), made this project possible.
References Cited


Chapter Seven

Student Feedback on Different Activity Formats Implemented in Upper-Division, Undergraduate Neurophysiology Lab Course at the University of Colorado Boulder

Diba Mani¹ and Adam Blanford²
Department of Integrative Physiology; Graduate Teacher Program

Introduction

This paper discusses the results of a Teaching-as-Research (TAR) project examining student feedback about different types of activity formats presented in an upper-division Neurophysiology lab at University of Colorado (CU) Boulder. Specifically, how students engage with lab activities and integrate their conceptual knowledge by using criteria developed for the Science Education Initiative (SEI).

The SEI was developed as a way to improve the quality of classroom education via three-part process: “(1) establish well-defined learning goals through faculty consensus, (2) create valid assessment tools for measuring attainment of these learning goals, and (3) create and use pedagogically effective materials and teaching approaches (Perkins et al. 2014:3).

To date, the SEI criteria have been applied to more than seven STEM departments at CU Boulder from 2005-2014 and affected approximately 10,000 students during that time. The SEI project has had a positive influence on undergraduate teaching and departmental culture through the establishment of consistent learning goals and teaching styles, which has led to an increase in both the frequency and quality of discussions about teaching and learning in the departments.

Despite its impacts on lecture-based courses, the SEI criteria have not been applied in STEM labs or other supplementary activities (e.g., lab or recitations held in parallel with the course lectures). After three years as a Teaching Assistant for a neurophysiology course, the PI
collaborated with TIGER and the Graduate Teacher Program (GTP) to apply aspects of the SEI project to the lab portion of the Neurophysiology course.

Prior to doing a TAR project, the PI did not have much experience with educational research aside from the typical Faculty Course Questionnaire (FCQ), which asks a standard battery of questions about instructor respect, intellectual challenge, and effort expended on the course each week. The TAR project showed what was possible for a graduate student teacher to do in the classroom to assess the quality of student learning.

TAR projects are small-scale classroom research projects administered by the Teaching Institute for Graduate Education Research (TIGER), as unit of the GTP at CU Boulder. These projects require the guidance of a faculty advisor/mentor, human subject research certification, and approval from the Institutional Review Board (IRB). This ensures proper steps have been taken to safeguard research data and participant information. The study was approved by the Institutional Review Board at the University of Colorado (Protocol #16-0701).

The PI did not expect to engage with the students as much as the SEI project had, nor was the present study as elaborate as the efforts of the SEI, but it was expected that one survey taken by students at the end of the term would suffice in getting preliminary feedback on the types of activities making up the lab syllabus, and that modifications would be made based on the fall 2016 course feedback for the spring 2017 course in a future study.

**Theoretical Background**

Students can learn through active or passive learning, as described by the Differentiated Overt Learning Activities (DOLA) structure (Chi 2009; Menekse et al. 2013). The academic literature is full of discussions about the pros and cons of “active” versus “passive” learning. “Passive” learning is typically associated with the traditional lecture-based classroom, where the teacher
lectures and students passively sit taking notes. Although this model is the standard for teaching in higher education, more researchers are calling for the incorporation of active learning techniques because active learning engages students in activities to strengthen their knowledge while improving their metacognitive skills (Menekse et. al. 2013; Prince 2004; Tanner 2012).

Catering the curricula to meet different learning styles (Angelo et al. 1998; Coffield et al. 2004; Gardner 1983; Honey and Mumford 1982; Kratzig and Arbuthnott 2006; Reynolds 1997) and giving students a sample of tasks completed in careers ranging from clinical settings as a neurologist to a graduate student and eventually a research professor have been implemented by others and shown to be beneficial (Donavan 1999, NRC Board on Life Sciences 2003). The influence of group versus independent activities have also been evaluated for educational merit (Crosling and Webb 2002; Duren and Cherrington 1992; Taylor 2011), which is particularly critical in teaching science (DeHaan 2005; Handelsman et al. 2007).

In this paper, we discuss how engaging students in classroom activities enables them to integrate their conceptual knowledge of a topic while expanding how they think about those concepts, inducing deeper understanding of the material and applying concepts in different settings (Tanner 2012). The lab is the perfect addition to the lecture aspect of the Neurophysiology course, which has a rather passive setting, aside from classroom questions scattered throughout the lecture. It is suggested that active learning is superior because students are literally participating in the activity (Prince 2004). Thus, in organizing a survey to acquire feedback on lab activities, we differentiated between hands-on, interactive activities, passive reading activities, and mid-level interactive activities, such as computer simulations and worksheets, which required some engagement but not as much as those that were hands-on and required physical movement and engagement.
Hands-on activities involve students reading a short document comprising background information on concepts and lab directions, which then leads into organizing the materials they need for the lab experiment, setting up equipment, and then implementing the experiment amongst themselves in small groups. Hands-on activities include wearing vision-altering goggles while walking or throwing balls at a target, using electrical stimulation to evoke muscle contractions, and evaluating muscle activation while maintain body posture and holding on to a bar on the wall. Passive reading activities involve being given a full manuscript on a relevant topic (i.e. prosthetics) to read individually and then work alone or in a small group to answer a series of questions pertaining to the paper. Mid-level interactive activities involve computer simulations, which require students to work around a computer in a small group to complete simulations on topics such as resting membrane potentials and equilibrium potentials. Other topics, such as action potentials and electrochemical gradients, are taught through a very old version of MS-DOS, which does not meet the technological advancement of other aspects of the course, given that it was last updated in 2000.

**Methods**

Thirty-six students enrolled in a Neurophysiology course in the fall 2016 semester provided consent though only 33 answered survey questions and were included in data analysis. The surveys were completed by students across 5 course sections. Students were provided a link via the course website to proceed for completion if they chose not to complete it during lab time.

The primary investigator (who is a lab coordinator, but with no direct lab setting interaction with the students) asked the teaching assistants in charge of the five lab sections to discuss the purpose of the project, administer informed consent by passing out copies of the form
to students, and to answer any questions the students might have. Exclusion criteria included students not enrolled in the course (auditing), taking the course for the second time, or not willing to participate, as it was predicted that these students would have a biased or comparative perspective (to the majority of students) regarding the course.

For those students who were eligible and agreed to participate, the opportunity to take the survey in the adjoining IPHY computer lab during class time was provided. Students were reminded that they had the option to decline to participate at any time with no penalties. The survey took less than 15 minutes to complete. Paper copies of the consent form was made available to students. Using the survey data, the primary investigator evaluated the level of student satisfaction, enjoyment, and applicability to make adjustments to the lab curriculum for spring 2017.

Originally, it was expected that, in addition to the survey, feedback would be compared against student grades and later to FCQs to determine what students thought of the course overall and to analyze their attendance and performance over the course of the semester. Specifically, these are lab attendance grades (to quantify individual lab attendance) and final overall course grade. In order to do this, students were asked to input their name and the last four digits of their student ID number for matching purposes. These survey questions were often not answered and not enough students were willing to share identifying information to associate their survey feedback with their course grades. Additionally, this request resulted in rather poor overall participation, and was suggested for removal for the spring 2017 term.

Survey questions (Appendix 1) included Likert-scale and open-ended questions designed to elicit responses on applicability, enjoyment/satisfaction, and motivations for taking the class, to provide background information and help eliminate any potentially confounding variables.
(i.e., Instructor enthusiasm or student motivations offsetting deficiencies in the lab). The Likert scale questions included several options (i.e. “How effective was each lab type in enhancing your understanding of lecture material?” – extremely effective, effective, moderately effective, somewhat effective, or not effective at all).

Results

More than half of the students (20 students, 60.6%) who participated in the survey answered that they found the hands-on investigative types of activities either “extremely effective” or “effective” in enhancing their understanding of lecture material. Nine students (27.27%) found computer simulation tutorials and reading literature to be more than moderately effective. Thirteen students (39.39%) found completing worksheets to be extremely effective or effective (Table 1).

<table>
<thead>
<tr>
<th>Question</th>
<th>Extremely Effective</th>
<th>Effective</th>
<th>Moderately Effective</th>
<th>Somewhat Effective</th>
<th>Not Effective at all</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative</td>
<td>27.33%</td>
<td>33.33%</td>
<td>11.15%</td>
<td>12.12%</td>
<td>12.12%</td>
<td>33</td>
</tr>
<tr>
<td>Computer simulation tutorial</td>
<td>6.06%</td>
<td>21.21%</td>
<td>30.30%</td>
<td>24.24%</td>
<td>18.18%</td>
<td>33</td>
</tr>
<tr>
<td>Reading literature</td>
<td>0.00%</td>
<td>27.27%</td>
<td>21.21%</td>
<td>33.33%</td>
<td>18.18%</td>
<td>33</td>
</tr>
<tr>
<td>Completing worksheets</td>
<td>15.15%</td>
<td>24.24%</td>
<td>30.30%</td>
<td>15.15%</td>
<td>15.15%</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 1 – Survey responses to the question, “How effective was each lab type in enhancing your understanding of lecture material?”

Typical comments from students who selected, “somewhat effective” or “not effective at all” for one or more lab categories showed a dislike for reading literature, that only “25% of the students read the actual paper during the lab period,” mostly due to a feeling of uselessness or
too dense to read through. Others commented that reading literature would be better done in private and not in the group lab setting, due to the complexity of the material and a lack of efficient use of time (i.e. “…not useful of the students’ time or the TA’s time. Articles/journal reviews are super detailed and dense with a lot of information that supplements topics slightly, but aren’t exactly helpful”).

Comments from students who selected, “extremely effective” or “effective” for one or more lab categories included an appreciation for the hands-on activities: “I am good at learning when I can work with my hands but also have a reading to go along with what we are learning”, “I liked how fun and intuitive the simulations and hands on were”, “I like doing hands on activities because I learn best from doing”, “I’m a visual learner and I also learn really well by doing things so the hands-on labs were most effective for me”.

Most students found the hands-on investigative (18 students, 56.25% voted “extremely effective” or “effective”) and the worksheets (16 students, 51.61%) to be the most effective activity types to assist in applying lecture concepts, again commenting that the interactive activities were superior to the “boring” and “unhelpful” literature.
The hands-on investigative material were by far the most enjoyed activity type (Figure 1), receiving votes from twenty-six students (80.25%) as enjoyed at least “a moderate amount”. The most favorite specific activities were the “partner-evoked contraction” and “drunk goggle plasticity” labs. Students also voiced that working in small groups was most helpful in enhancing learning than working with one partner or alone (Table 2).
<table>
<thead>
<tr>
<th>Number</th>
<th>Answer</th>
<th>Percentage</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Working in small groups</td>
<td>64.52%</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>Working with a partner</td>
<td>9.68%</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Working independently</td>
<td>9.68%</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Combination of independent and with others</td>
<td>16.13%</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>No preference</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>100%</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2 – Survey response to the question, “which type of lab collaboration was most helpful in enhancing your learning?”

Overall, students appreciate the graded lab quizzes, which were the key determinant in encouraging students to read lab materials in advance or to prepare for lab activities. Most found grading to be “very fair” (25 students, 78.13%) or “moderately fair” (7 students, 21.88%). Changes suggested included a few requests for fewer quizzes and more overall course weight on the lab portion of the course.

When asked which specific labs they would suggest retaining for future semesters, most selections were for interactive activities such as the partner-evoked contractions and drunk goggle plasticity labs (Figure 2). Twenty-two students agreed that exam review should be kept for future semesters. Only 4-6 students suggested retaining review documents, hence implying that reading documents in lab is not viewed favorably by students. All other activities inquired about, including computer simulations, conceptual demonstrations, and on-hands activities such as the conduction velocity and EMG with weights labs, were suggested for retention by 10-18 students.
Figure 2 – Survey response to the question, “which specific labs would you recommend retaining for future semesters? Select all that apply.”

Analysis

The interactive, hands-on activities received the best feedback from the most students. Students found that activities that expanded on concepts taught in lecture were the most useful and that the engagement of teaching assistants was appreciated, vocalizing their own learning styles in the free response portion of the survey. Predominantly, the literature reviews were considered an ineffective use of time and not conductive to elaborating concepts learned in lecture. However, based on voting results, a variety of activity styles was appreciated, as commented by a few students: “all of the categories were effective because they brought the lecture material into use that was easier to understand or to see how they were used in real world applications. They were
also great study material for exams” and, “the labs are fun and then seem to work most of the time. I like being able to see what I have learned in action”.

While it was expected that more students would appreciate the independent reading literature component of the lab, survey feedback demonstrates that students expect and appreciate the lab component of the course to target group work and hands-on engagement. Given that a few students did select reading literature as an effective means for learning course material, it is suggested that there still be some preservation of literature review in future lab course syllabi.

Conclusion

The organization of lab activities based on learning styles similar to those described in the DOLA framework was useful in revealing that active learning is preferred in the lab portion of the undergraduate Neurophysiology course. Engaging students in group interaction to expand on concepts taught in lecture is important for retention of course concepts. The passive activities are considered dull and a waste of time, and should be removed from future lab syllabi.

Future investigation should implement these changes and evaluate student feedback following these changes. It may be useful to create more active learning lab activities to replace those that are repetitive and non-engaging with students.
References Cited


Appendix 1: Survey administered to students via Qualtrics (2016, Provo, UT, USA).

IPHY 4720 Neurophysiology Lab Activity Survey

Q1 Title of research study: Lab Activities to Supplement CU Boulder IPHY 4720 Neurophysiology Lecture FA 16-SP 17

Investigator: Diba Mani, M.S.

Why am I being invited to take part in a research study?
We invite you to take part in a research study because you are enrolled in IPHY 4720 Neurophysiology for the Fall 2016 or Spring 2017 terms.

What should I know about a research study?
· Someone will explain this research study to you.
· Whether or not you take part is up to you.
· You can choose not to take part.
· You can agree to take part and later change your mind.
· Your decision will not be held against you.
· You can ask all the questions you want before you decide.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt you, talk to Diba Mani via diba.mani@colorado.edu. This research has been reviewed and approved by an Institutional Review Board (“IRB”). You may talk to them at (303) 735-3702 or irbadmin@colorado.edu if:
· Your questions, concerns, or complaints are not being answered by the research team.
· You cannot reach the research team.
· You want to talk to someone besides the research team.
· You have questions about your rights as a research subject.
· You want to get information or provide input about this research.

Why is this research being done?
We are conducting a comparative study of fall 2016 and spring 2017 IPHY 4720 lab sections using surveys, student grade data, and Faculty Course Questionnaires (FCQs) to whether your experiences in the class have been consistent with guidelines laid out in the Science Education Initiative (SEI), which has been used in the IPHY department to evaluate teaching/learning methods in lectures, as well as if you feel your experiences have been relevant to your career path and learning objectives. These data will help us determine the extent to which changes are necessary and help the department evaluate its future commitments to the SEI guidelines.

How long will the research last?
Your active participation will last approximately fifteen minutes, for the duration of completing the survey. We expect that your information (e.g., survey data, grade data) will be utilized in this research study through July 2017, at the latest, when data analysis on the surveys will be completed.
How many people will be studied?
We expect up to 140 people will be in this research study during the fall 2016 and spring 2017 semesters

What happens if I say yes, I want to be in this research?
You will be asked to sign an informed consent form during the class. You will complete the survey during an IPHY lab in the last two weeks of class (providing all activities are complete in the IPHY computer lab located in CLRE 111. We are also asking for access to student grades to look at performance in the lab during the semester you are enrolled. Specifically, these are lab attendance grades (to quantify individual lab attendance) and final overall course grade. These data will be collected at the end of the semester, matched against your survey, and then anonymized prior to analysis. FCQ results will also be collected when they become publicly available.

What happens if I do not want to be in this research?
You can leave the research at any time and it will not be held against you.

What happens if I say yes, but I change my mind later?
You can leave the research at any time it will not be held against you.

Is there any way being in this study could be bad for me? No risks are expected.

Will being in this study help me in any way?
We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include improving the course labs for future terms. There will absolutely be no impact on your grade.

What happens to the information collected for the research?
All identifying information will be anonymized prior to analysis. Data collected will be password-protected and only accessible by the Primary Investigator.

What else do I need to know?
This research is being funded by the University of Colorado Boulder Graduate Teacher Program’s initiative, Teaching as Research (TAR).

By clicking yes, you indicate that you are over the age of 18 and you are consenting to participate in the study.
☐ Yes (1)
☐ No (2)

If No Is Selected, Then Skip To End of Survey

Q30 Please type your first name.
Q31 Please type the last four digits of your student ID number (XXXX).

Q2 Is this your first time taking IPHY 4720?
- Yes (1)
- No (2)

If No Is Selected, Then Skip To End of Survey

Q3 We've broken down the lab activities in IPHY 4720 into four distinct categories: (a) Hands-on investigative a. Partner-evoked contractions w/ electrical stimulation b. Conduction velocity c. EMG with weights d. Reflex demonstration e. Adaptation w/ goggles, balls, pennies, and baskets (b) Computer simulation tutorial a. Action potential tutorial b. GHK, Nernst Equation application (c) Reading literature a. Plasticity and rehabilitation review document b. Neuroprosthesis and grasping review document (d) Completing worksheets a. SI units, Ohm's Law, conversions, signal transduction, logs/exponents

How effective was each lab type in enhancing your understanding of lecture material?

<table>
<thead>
<tr>
<th></th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Q4 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Answer If We've broken down the lab activities in IPHY 4720 into three distinct categories: (a) hands-on... - Somewhat Effective Is Selected Or We've broken down the lab activities in IPHY 4720 into three distinct categories: (a) hands-on... - Not effective at all Is Selected

Q25 You answered "Extremely effective" or "Effective" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?
Q5 How effective was each lab type in applying lecture concepts through the activities?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (5)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Answer If How effective was each lab type in applying concepts through the activities? - Extremely effective Is Selected

Or How effective was each lab type in applying concepts through the activities? - Very effective Is Selected

Q26 You answered "Extremely effective" or "Effective" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Answer If How effective was each lab type in applying concepts through the activities? - Somewhat Effective Is Selected

Or How effective was each lab type in applying concepts through the activities? - Not effective at all Is Selected

Q6 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Q9 How much did you enjoy completing activities within each lab category?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>A great deal (1)</th>
<th>A lot (2)</th>
<th>A moderate amount (3)</th>
<th>A little (4)</th>
<th>Not at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Q10 Can you provide an example of what activity/activities you particularly enjoyed and why?
Q11 Which type of lab collaboration was most helpful in enhancing your learning?
- Working in small groups (1)
- Working with a partner (2)
- Working independently (3)
- Combination of independent & with others (4)
- No preference (5)

Q29 Did the lab quizzes encourage you to read lab materials in advance or prepare for lab activities?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- Not at all (5)

Q32 How fair was the grading format for the lab?
- Very fair (1)
- Moderately fair (2)
- Not fair (3)

Q33 How could the grading be improved?

Q13 How much did each lab type help prepare you for your education (i.e. graduate school) or professional goals?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>A great deal (1)</th>
<th>A lot (2)</th>
<th>A moderate amount (3)</th>
<th>A little (4)</th>
<th>None at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
<td>❍</td>
</tr>
</tbody>
</table>

Q33 Which type of lab type would you like to see more of?
- Hands-on investigative (1)
- Computer simulation tutorial (2)
- Reading literature (3)
- Completing worksheets (4)
- None or other (5) ____________________
Q31 How much has the lab has increased your desire to continue learning about this material?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- None at all (5)

Q15 Which specific labs would you recommend retaining for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Conduction velocity (2)
- EMG with weights (3)
- Reflex demonstration (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Exam reviews (10)

Q16 Which specific labs would you recommend removing for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Conduction velocity (2)
- EMG with weights (3)
- Reflex demonstration (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Exam reviews (10)

Q17 How much do you think the experiences acquired in the lab will impact your future career or educational goals?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- Not at all (5)
Q18 What are your future goals (career/education)?

- MD/DO (1)
- PT (2)
- Nursing (3)
- Education (4)
- Graduate school (5)
- Other (please specify) (6) ____________________

Q19 Why did you enroll in this course (i.e. required course, interested in topic, fit schedule best)? Select all that apply.

- Required for degree (1)
- General interest (2)
- Prepare for future goals (3)
- Fit into schedule best (4)
- Other (please specify) (5) ____________________

Q20 Please provide any additional feedback on the lab portion of this course.
Chapter Eight

Student Feedback on Different Activity Formats Implemented in Upper-Division, Undergraduate Neurophysiology Lab Course at the University of Colorado Boulder, Following Pilot Study Modifications from the Previous Semester

Diba Mani
Department of Integrative Physiology, CU Boulder

Introduction

This paper discusses the second iteration of the TAR project in the IPHY Neurophysiology course after making changes suggested in fall 2016. Labs and recitations associated with upper-division IPHY courses typically enhance student understanding and retention of lecture material by expanding on topics discussed in lecture. To date, the Integrative Physiology (IPHY) department has spent time implementing Science Education Initiative (SEI) activities in a number of upper-division courses with an aim toward improvement (Perkins et al. 2015). The IPHY department participated in CU Boulder SEI during the years 2007-2012. An example of an objective reached in 2012 was the addition of 3 inquiry-based lab activities for the Physiology course labs (Perkins et al. 2012).

The direction that SEI has provided has helped adjust IPHY course connectivity between not only labs/recitations and lectures but also enabled congruency between lecture material, homework assignments, and supplementary activities. These modifications have led to improved student performance and an enhanced experience. It was found that connecting the lecture concepts to real-world applications (clinical and research) are interesting to students and better enable student engagement with the material and professor.

Although SEI initiatives show a clear benefit to the IPHY lecture-based courses, the Neurophysiology labs have not received similar attention. As a result, we intend to evaluate
student experiences without significant changes instituted, then make changes based on feedback and learning improvement criteria provided by programs like SEI, and finally determine whether those changes have helped improve student experiences and learning. That would help the department adapt labs to SEI standards more efficiently with a comparative dataset. Many other courses in the IPHY department (and others) have 2+ hour weekly labs and recitations; we need to use these hours as effectively as possible to meet the best standards for learning. Starting with the IPHY labs with the intention of progressing to other labs and recitations is the most logical format for an eventual expansion of this objective.

Thus far, these modifications have reached the IPHY course lecture but not the lab, predominantly because the lab has the history of being led and designed by the graduate students, who typically do not have a vested interest in lab enhancement even when they have the experience to make changes. Examples of negative feedback from students in this course includes the use of outdated technology, such as the use of Windows 98 for a computer application to study action potentials, and non-practical lesson plans, such as lab assignments running too long one week and not long enough the following, that are often used in the course. As many of these undergraduates are IPHY majors who plan to utilize their education in an applied field, taking full advantage of the labs to learn the specialized techniques and skills that these potential careers entail is ideal. Additionally, it has been observed that, pending the professor and emphasis of lecture material, be it research-based or clinical-based, the lab activities should be modified to best meet the intention of the two-hour per week time allocated.

Based on anecdotal evidence, the PI made changes to Neurophysiology labs during the fall 2016 semester when the Teaching As Research (TAR) project was first implemented – for example, the motor unit demonstration lab was been omitted for the fall term. Other adjustments
The purpose of the current study was to evaluate how specific changes to lab activity and structure would impact student learning and enjoyment of the lab portion of the course. These changes were made at the suggestion of students who completed the survey at the end of the previous term. A secondary aim was to make sure that different styles of learning, be it from hands-on interactive activities or from reading literature on one’s own, were still respected in the course syllabi of activities. It was expected that the changes would improve student perspectives on the lab but that some modifications would still not be enjoyed by students but were necessary for teaching an important concept (i.e. reducing but keeping some literature review activities).

Theoretical Background

The importance of catering curricula to meet the needs of students representing all learning styles is vital (Angelo et al. 1998; Coffield et al. 2004; Gardner 1983; Honey and Mumford 1982; Kratzig and Arbuthnott 2006; Reynolds 1997), as has relating course materials to future professions been proven beneficial (Donavan 1999; NRC Board on Life Sciences 2003). To do this successfully, we implement active and passive learning strategies (Chi 2009), as described by the Differentiated Overt Learning Activities (DOLA) structure. “Active” learning involves students interacting in activities in a “hands-on” method, strengthening knowledge and
improving metacognitive skills (Prince 2004; Tanner 2012). “Passive” learning is most typical in lectures, and involves receiving information, and can also be considered the means of learning when reading an article (Chi 2009).

Given that the lecture gives way to passive learning more than active, the lab component of the Neurophysiology course is integral in promoting active learning and enabling group engagement (Crosling and Webb 2002; Duren and Cherrington 1992; Taylor 2011), which is particularly critical in teaching science (DeHaan 2005; Handelsman et al. 2007).

In this paper, we build on the feedback provided by students about the lab portion of the Neurophysiology course at the end of fall 2016 to further enhance learning and effectiveness of the lab during the spring 2017 term, which encompasses nearly two hours of time per week for the 15-week length of the course. The engagement of different learning styles and further group collaboration was expected to increase positive feedback from the survey acquired from the students at the end of the spring 2017 term.

The changes to the spring term were relevant to the particular activities implemented during the lab and the course grading scheme. Specifically, rather than 3 full manuscripts to read and answer questions about, only 1 was included in the spring term. These openings in activity schedule were replaced with new hands-on activities such as the “drunk goggles” lab, which evaluates concepts of motor adaptation and sensory feedback, and the postural control lab, which considers muscle activation in different parts of the body in response to an external stimulus. Student feedback from the fall did not reveal a huge emphasis for change in the grading scheme, so we removed all graded portions of the lab for the spring to compare if and how attendance, course engagement, and student learning changed from one term to the next.
Methods

Fifty-seven students enrolled in Neurophysiology in the Spring 2017 semester provided consent though only fifty-four answered survey questions and were included in data analysis, as the other three selected that they had taken a Neurophysiology course previously. The surveys were completed by students across 5 course sections. Students were provided a link via the course website to proceed for completion if they chose not to complete it during lab time. The study was approved by the Institutional Review Board at the University of Colorado Boulder (Protocol #16-0701).

The primary investigator asked the teaching assistants in charge of the five lab sections to discuss the purpose of the project, administer informed consent by passing out copies of the form to students, and to answer any questions the students might have. Additionally, students were offered one point of extra credit for their participation, and questions asking for student names were removed from the survey, based on the results from the fall 2016 survey. Exclusion criteria included students not enrolled in the course (auditing), taking the course for the second time, or not willing to participate, as it was predicted that these students would have a biased or comparable perspective (to the majority of students) regarding the course.

For those students who were eligible and agreed to participate, the opportunity to take the survey in the adjoining IPHY computer lab during class time was provided. Students were reminded that they had the option to decline to participate at any time with no penalties. The survey took less than 15 minutes to complete. Paper copies of the consent form was made available to students. The results from the survey taken in the spring was compared to those acquired in the fall.
Survey questions (Appendix 1, Appendix 2) included Likert-scale and open-ended questions designed to elicit responses on applicability, enjoyment/satisfaction, and motivations for taking the class, which will provide background information and help eliminate any potentially confounding variables (i.e., instructor enthusiasm or student motivations offsetting deficiencies in the lab). The Likert scale questions included several options (i.e. “How effective was each lab type in enhancing your understanding of lecture material?” – extremely effective, effective, moderately effective, somewhat effective, or not effective at all). These questions were the same ones asked during the fall 2016 term, with the appropriate modifications made to the specific lab activities implemented in the new term.

**Results**

Compared to 60.6% of students in the fall semester responding that they found the hands-on investigative types of activities either “extremely effective” or “effective” in enhancing their understanding of lecture material, 72.23% responded as such at the end of the spring semester. Compared to nine students (27.27%) in the fall, nineteen students (35.19%) found computer simulation tutorials to be more than moderately effective; eleven (20.37%) found reading literature more than moderately effective in the spring. Thirteen students (39.39%) found completing worksheets to be extremely effective or effective (Table 1).

<table>
<thead>
<tr>
<th>Choices</th>
<th>Extremely</th>
<th>Effective</th>
<th>Moderately</th>
<th>Somewhat</th>
<th>Not</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115</td>
</tr>
</tbody>
</table>

115
Typical comments from students who selected, “somewhat effective” or “not effective at all” for one or more lab categories still showed a dislike for reading literature, describing them as “dense”, “bit tedious and it’s hard to [focus on]”, and “so complex that it was hard to understand the main ideas/how they related to the course lecture”. Comments from students who selected, “extremely effective” or “effective” for one or more lab categories included an appreciation for the hands-on activities complementing lecture material well, and that they encouraged students to grasp concepts taught in lecture more fully. Students expressed an appreciation for worksheets as promoting the independent learning: “the worksheets helped because it was more or less independent work. I do better with this” and “for me personally, worksheets with practice problems are the most helpful way for me to understand concepts”. Other comments appreciated the partner-evoked contractions, the drunk goggle plasticity lab, and stretching labs most profusely.

In the fall (Figure 1A), most students found the hands-on investigative (18 students, 56.25%) voted “extremely effective” or “effective” and the worksheets (16 students, 51.61%) to be the most effective activity types to assist in applying lecture concepts, which was reiterated in the spring (Figure 1B) – 34 students (65.39%) found the hands-on investigative more than

<table>
<thead>
<tr>
<th>Hands-on investigative</th>
<th>Effective (FA 16-SP 17)</th>
<th>Effective</th>
<th>Effective</th>
<th>Effective at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.33-25.93% 14</td>
<td>33.33-46.30% 25</td>
<td>15.15-12.96% 7</td>
<td>12.12-14.81% 8</td>
<td>12.12-0.00% 0</td>
</tr>
<tr>
<td>6.06-5.56% 3</td>
<td>21.21-29.63% 16</td>
<td>30.30-38.89% 21</td>
<td>24.24-22.22% 12</td>
<td>18.18-3.70% 2</td>
</tr>
<tr>
<td>0.00-5.56% 3</td>
<td>27.27-14.81% 8</td>
<td>21.21-38.89% 21</td>
<td>33.33-27.78% 15</td>
<td>18.18-12.96% 7</td>
</tr>
<tr>
<td>15.15-7.41% 4</td>
<td>24.24-42.59% 23</td>
<td>30.30-33.33% 118</td>
<td>15.15-11.11% 6</td>
<td>15.15-5.56% 3</td>
</tr>
</tbody>
</table>

Table 1 – Survey responses to the question, “How effective was each lab type in enhancing your understanding of lecture material?”
moderately effective; 26 students (50.98%) found the worksheets also more than moderately effective.
In the fall, hands-on investigative material received votes from twenty-six students (80.25%) as enjoyed at least “a moderate amount”; this was consistent with results in the spring, with 71.67% considering the hands-on material as enjoyed “a great deal” or “a lot.” The most favorite specific activities were the “partner-evoked contraction” and “drunk goggle plasticity” labs in the fall and the spring, with additional positive comments on the stretching and locomotion labs. Students also voiced that working in small groups was most helpful in
enhancing learning than working with one partner or alone in both the fall (Table 2) and the spring.

<table>
<thead>
<tr>
<th>Number</th>
<th>Answer</th>
<th>Percentage (FA 16)</th>
<th>Count (FA 16)</th>
<th>Percentage (SP 17)</th>
<th>Count (SP 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Working in small groups</td>
<td>64.52%</td>
<td>20</td>
<td>81.13%</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>Working with a partner</td>
<td>9.68%</td>
<td>3</td>
<td>1.89%</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Working independently</td>
<td>9.68%</td>
<td>3</td>
<td>3.77%</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Combination of independent and with others</td>
<td>16.13%</td>
<td>5</td>
<td>13.21%</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>No preference</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100%</td>
<td>31</td>
<td>100%</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 2 – Survey response to the question, “which type of lab collaboration was most helpful in enhancing your learning?” in the fall (A) and the spring (B).

One change made between fall and spring is that spring labs were no longer graded in the areas of participation and effort. Most found grading to be “very fair” (25 students, 78.13%) or “moderately fair” (7 students, 21.88%) in the fall term, as well as the spring (43 students, 81.13%), with a pretty similar breakdown between the levels of fairness. There is no significant difference and thus it is considered that the inclusion of a grade in the lab component of the course does not result in a change in student effort. The only consistent comment, beside no qualms on the “no grading” of the lab in the spring, was the addition of “participation” points to motivate more students to attend.

When asked which specific labs they would suggest retaining for future semesters, most selections were for interactive activities such as the partner-evoked contraction and drunk goggle plasticity labs in the fall (Figure 2A); the partner-evoked contractions and stretching labs were key for the spring term students (Figure 2B). Activities that involved reading literature were vehemently suggested for removal.
Figure 2 – Survey response to the question, “which specific labs would you recommend retaining for future semesters? Select all that apply.” for the fall (A) and spring (B) terms.

Analysis

The improved perspective on effectiveness on computer simulations and reading literature is likely due to the fact that there were fewer in the spring than the fall. There were also more hands-on interactive activities, including the addition of a brand-new lab activity created during winter break between terms. The inclusion of the reading material in the spring was to make up for the lack of lengthy pre-lab descriptions of activities and to demonstrate how researchers (including graduate students) review manuscripts based on research studies. This attempt to parallel future ambitions of students was seemingly poorly done and could be improved but should not be omitted, given the predominantly negative response students had toward reading literature over participating in hands-on activities. Additionally, mentioning that the readings are not for retaining knowledge but rather to demonstrate how manuscripts and research projects are completed is vital. Modifications for future semesters were immediately considered following the receipt of this data.

Due to overwhelming appreciation of the hands-on interactive labs, especially the partner-evoked contractions and the drunk goggles plasticity, it is recommended to keep this labs for future terms. Additionally, the locomotion and stretching labs went well and should be included. The positive feedback on the devices used for the electrical stimulation resulted in the acquisition of similar labs to be used for the upcoming term (summer 2017).

There was considerably more engagement in feedback in the spring semester than the fall. The comments provided by students in the spring were more in-depth and constructive.
Likely, this is due to the one point extra credit rewarded, as well as the feedback the primary investigator gave the other TAs in acquiring informed consent about the study.

Modifying the course grading scheme in the spring to include participation credit for the lab component of the course will be considered, given that attendance was lower in the spring than the fall. However, given that there were no complaints between semesters, changes to the grading scheme are considered secondary.

**Conclusion**

The modifications made to the spring 2017 term based on the fall 2016 term survey were effective in improving student feedback and learning. Again, students expressed an appreciation for different learning styles but preferred on-hands interactive activities to complement the material learning in the more passive lecture component of the course. Still, there are changes that should be made to further improve the course. As shown over several years of work by the SEI project, improving course learning methods and syllabi takes several semesters but is certainly possible and a worthwhile effort.

Future studies should focus on improving the literature component of the labs, such as writing shorter, more concise reviews of only course-related concepts and research methods. Unfortunately, these would not be reflective of the actual types of readings graduate students and other researchers review. Additionally, perfecting the hands-on interactive labs so that their successful implementation is consistent across course sections and student groups. The suggestion is that training each TA with a full review of the labs prior to completing the activity with the students may be helpful to assure a successful lab activity.
References Cited


Appendix 1

Survey administered to students via Qualtrics (2016, Provo, UT, USA) in Fall 2016.

IPHY 4720 Neurophysiology Lab Activity Survey

Q1 Title of research study: Lab Activities to Supplement CU Boulder IPHY 4720

Neurophysiology Lecture FA 16-SP 17
Investigator: Diba Mani, M.S.

Why am I being invited to take part in a research study?
We invite you to take part in a research study because you are enrolled in IPHY 4720 Neurophysiology for the Fall 2016 or Spring 2017 terms.

What should I know about a research study?
· Someone will explain this research study to you.
· Whether or not you take part is up to you.
· You can choose not to take part.
· You can agree to take part and later change your mind.
· Your decision will not be held against you.
· You can ask all the questions you want before you decide.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt you, talk to Diba Mani via diba.manii@colorado.edu. This research has been reviewed and approved by an Institutional Review Board (“IRB”). You may talk to them at (303) 735-3702 or irbadmin@colorado.edu if:
· Your questions, concerns, or complaints are not being answered by the research team.
· You cannot reach the research team.
· You want to talk to someone besides the research team.
· You have questions about your rights as a research subject.
· You want to get information or provide input about this research.

Why is this research being done?
We are conducting a comparative study of fall 2016 and spring 2017 IPHY 4720 lab sections using surveys, student grade data, and Faculty Course Questionnaires (FCQs) to whether your experiences in the class have been consistent with guidelines laid out in the Science Education Initiative (SEI), which has been used in the IPHY department to evaluate teaching/learning methods in lectures, as well as if you feel your experiences have been relevant to your career path and learning objectives. These data will help us determine the extent to which changes are necessary and help the department evaluate its future commitments to the SEI guidelines.

How long will the research last?
Your active participation will last approximately fifteen minutes, for the duration of completing the survey. We expect that your information (e.g., survey data, grade data) will be utilized in this research study through July 2017, at the latest, when data analysis on the surveys will be completed.
How many people will be studied?
We expect up to 140 people will be in this research study during the fall 2016 and spring 2017 semesters.

What happens if I say yes, I want to be in this research?
You will be asked to sign an informed consent form during the class. You will complete the survey during an IPHY lab in the last two weeks of class (providing all activities are complete in the IPHY computer lab located in CLRE 111. We are also asking for access to student grades to look at performance in the lab during the semester you are enrolled. Specifically, these are lab attendance grades (to quantify individual lab attendance) and final overall course grade. These data will be collected at the end of the semester, matched against your survey, and then anonymized prior to analysis. FCQ results will also be collected when they become publicly available.

What happens if I do not want to be in this research?
You can leave the research at any time and it will not be held against you.

What happens if I say yes, but I change my mind later?
You can leave the research at any time it will not be held against you.

Is there any way being in this study could be bad for me?
No risks are expected.

Will being in this study help me in any way?
We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include improving the course labs for future terms. There will absolutely be no impact on your grade.

What happens to the information collected for the research?
All identifying information will be anonymized prior to analysis. Data collected will be password-protected and only accessible by the Primary Investigator.

What else do I need to know?
This research is being funded by the University of Colorado Boulder Graduate Teacher Program’s initiative, Teaching as Research (TAR).

By clicking yes, you indicate that you are over the age of 18 and you are consenting to participate in the study.
☐ Yes (1)
☐ No (2)

If No Is Selected, Then Skip To End of Survey

Q30 Please type your first name.
Q31 Please type the last four digits of your student ID number (XXXX).

Q2 Is this your first time taking IPHY 4720?
   ☐ Yes (1)
   ☐ No (2)

If No Is Selected, Then Skip To End of Survey

Q3 We've broken down the lab activities in IPHY 4720 into four distinct categories:

(a) Hands-on investigative
   a. Partner-evoked contractions with electrical stimulation
   b. Conduction velocity
   c. EMG with weights
   d. Reflex demonstration
   e. Adaptation with goggles, balls, pennies, and baskets

(b) Computer simulation tutorial
   a. Action potential tutorial
   b. GHK, Nernst Equation application

(c) Reading literature
   a. Plasticity and rehabilitation review document
   b. Neuroprosthesis and grasping review document

(d) Completing worksheets
   a. SI units, Ohm's Law, conversions, signal transduction, logs/exponents

How effective was each lab type in enhancing your understanding of lecture material?

<table>
<thead>
<tr>
<th></th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Q4 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Q5 How effective was each lab type in applying lecture concepts through the activities?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completing worksheets (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q6 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?
Q9 How much did you enjoy completing activities within each lab category?

<table>
<thead>
<tr>
<th>Activity</th>
<th>A great deal (1)</th>
<th>A lot (2)</th>
<th>A moderate amount (3)</th>
<th>A little (4)</th>
<th>Not at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Q10 Can you provide an example of what activity/activities you particularly enjoyed and why?

Q11 Which type of lab collaboration was most helpful in enhancing your learning?
- Working in small groups (1)
- Working with a partner (2)
- Working independently (3)
- Combination of independent & with others (4)
- No preference (5)

Q29 Did the lab quizzes encourage you to read lab materials in advance or prepare for lab activities?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- Not at all (5)

Q32 How fair was the grading format for the lab?
- Very fair (1)
- Moderately fair (2)
- Not fair (3)

Q33 How could the grading be improved?
Q13 How much did each lab type help prepare you for your education (i.e. graduate school) or professional goals?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>A great deal (1)</th>
<th>A lot (2)</th>
<th>A moderate amount (3)</th>
<th>A little (4)</th>
<th>None at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Q33 Which type of lab type would you like to see more of?
- Hands-on investigative (1)
- Computer simulation tutorial (2)
- Reading literature (3)
- Completing worksheets (4)
- None or other (5) ____________________

Q31 How much has the lab has increased your desire to continue learning about this material?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- None at all (5)

Q15 Which specific labs would you recommend retaining for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Conduction velocity (2)
- EMG with weights (3)
- Reflex demonstration (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Exam reviews (10)
Q16 Which specific labs would you recommend removing for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Conduction velocity (2)
- EMG with weights (3)
- Reflex demonstration (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Exam reviews (10)

Q17 How much do you think the experiences acquired in the lab will impact your future career or educational goals?
- A great deal (1)
- A lot (2)
- A moderate amount (3)
- A little (4)
- Not at all (5)

Q18 What are your future goals (career/education)?
- MD/DO (1)
- PT (2)
- Nursing (3)
- Education (4)
- Graduate school (5)
- Other (please specify) (6) ____________________

Q19 Why did you enroll in this course (i.e. required course, interested in topic, fit schedule best)? Select all that apply.
- Required for degree (1)
- General interest (2)
- Prepare for future goals (3)
- Fit into schedule best (4)
- Other (please specify) (5) ____________________

Q20 Please provide any additional feedback on the lab portion of this course.
Appendix 2


IPHY 4720 Neurophysiology Lab Activity Survey - SP 17
Title: Lab Activities to Supplement CU Boulder IPHY 4720 Neurophysiology Lecture SP 17
Investigator: Diba Mani, M.S.

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are enrolled in IPHY 4720 Neurophysiology for the Fall 2016 or Spring 2017 terms.

What should I know about a research study?

- Someone will explain this research study to you.
- Whether or not you take part is up to you.
- You can choose not to take part.
- You can agree to take part and later change your mind.
- Your decision will not be held against you.
- You can ask all the questions you want before you decide.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to Diba Mani via dibanani@colorado.edu. This research has been reviewed and approved by an Institutional Review Board (“IRB”). You may talk to them at (303) 735-3702 or irbadmin@colorado.edu if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research subject.
- You want to get information or provide input about this research.

Why is this research being done?

Similar surveys implemented previously have shown that student feedback is imperative in maximizing the benefit of the labs and lectures for IPHY courses. Changes implemented from anecdotal evidence and informal student feedback will be modified as necessary based on your comments for future semesters. This will improve the effectiveness of the labs in enhancing lecture material.

How long will the research last?

We expect that your information will be utilized in this research study through July 2017, at the latest, when data analysis on the surveys will be completed. Your active participation will last approximately fifteen minutes, for the duration of completing the survey.

How many people will be studied?

We expect up to 150 people will be in this research study.

What happens if I say yes, I want to be in this research?

You will be asked to complete the survey either during an IPHY lab in the last two weeks of class (providing all activities are complete) or one your own using a survey link sent through the class email.
We are also asking for access to student grades to look at performance in the lab during the semester you are enrolled. These data will be anonymized prior to any analysis.

**What happens if I do not want to be in this research?**
You can leave the research at any time and it will not be held against you.

**What happens if I say yes, but I change my mind later?** You can leave the research at any time it will not be held against you.

**Is there any way being in this study could be bad for me?**
No risks are expected.

**Will being in this study help me in any way?**
We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include improving the course labs for future terms. There will absolutely be no impact on your grade.

**What happens to the information collected for the research?** All identifying information will be anonymized prior to analysis. Data collected will be password-protected and only accessible by the Primary Investigator.

**What else do I need to know?**
This research is being funded by the University of Colorado Boulder Graduate Teacher Program’s initiative, Teaching as Research (TAR). By clicking yes, you indicate that you are over the age of 18 and you are consenting to participate in the study.

Q31 If you want to provide feedback on grades (anonymous), please type the last four digits of your student ID number (XXXX).

Q2 Is this your first time taking IPHY 4720?
Yes (1)
No (2)

If No Is Selected, Then Skip To End of Survey
Q3 We've broken down the lab activities in IPHY 4720 into four distinct categories:

(a) Hands-on investigative
   a. EMG with weights
   b. Partner-evoked contractions w/ electrical stimulation
   c. Motor unit demonstration in Enoka Lab
   d. M-wave, H-reflex demonstration.
   e. Postural control feedforward lab
   f. Adaptation w/ goggles, balls, pennies, and baskets (plasticity)
   g. Reflexes and stretching: tendon tap, balls, and partnered techniques.
   Kinesthesia vibrator lab
   i. Locomotion

(b) Computer simulation tutorial
   a. Action potential tutorial
   b. GHK, Nernst Equation (equilibrium, resting membrane potential) application

(c) Reading literature
   a. Magnusson et al. (1994) paper and discussion

(d) Completing worksheets
   a. Units of measurements, SI units, conversions

How effective was each lab type in enhancing your understanding of lecture material?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If we've broken down the lab activities in IPHY 4720 into four distinct categories:
(a) Hands-on investigative  
  a. EMG with weights  
  b. &nb...
- Somewhat effective Is Selected
Or we've broken down the lab activities in IPHY 4720 into four distinct categories:
(a) Hands-on investigative  
  a. EMG with weights  
  b. &nb...
- Not effective at all Is Selected
Q4 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

If we've broken down the lab activities in IPHY 4720 into four distinct categories:
(a) Hands-on investigative  
  a. EMG with weights  
  b. &nb...
- Extremely effective Is Selected
Or we've broken down the lab activities in IPHY 4720 into four distinct categories:
(a) Hands-on investigative  
  a. EMG with weights  
  b. &nb...
- Effective Is Selected
Q25 You answered "Extremely effective" or "Effective" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Q5 How effective was each lab type in applying lecture concepts through the activities?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
<td>☺</td>
</tr>
</tbody>
</table>

If How effective was each lab type in applying concepts through the activities? - Extremely effective Is Selected
Or How effective was each lab type in applying concepts through the activities? - Very effective Is Selected
Q26 You answered "Extremely effective" or "Effective" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?
Q6 You answered "Somewhat effective" or "Not effective at all" for one or more lab categories. Would you please provide an example explaining why you selected your particular answer?

Q9 How much did you enjoy completing activities within each lab category?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Q10 Can you provide an example of what activity/activities you particularly enjoyed and why?

Q11 Which type of lab collaboration was most helpful in enhancing your learning?

Working in small groups (1)
Working with a partner (2)
Working independently (3)
Combination of independent & with others (4)
No preference (5)

Q29 How much would the addition of lab quizzes encourage you to read lab materials in advance or prepare for lab activities?

A great deal (1)
A lot (2)
A moderate amount (3)
A little (4)
Not at all (5)
Q32 How fair was the grading format for the lab (no grading)?
Very fair (1)
Moderately fair (2)
Not fair (3)

Q33 How could the "grading" be improved?

Q13 How much did each lab type help prepare you for your education (i.e. graduate school) or professional goals?

<table>
<thead>
<tr>
<th>Lab Type</th>
<th>Extremely effective (1)</th>
<th>Effective (2)</th>
<th>Moderately effective (3)</th>
<th>Somewhat Effective (4)</th>
<th>Not effective at all (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on investigative (1)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Computer simulation tutorial (2)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reading literature (3)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Completing worksheets (4)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Q33 Which type of lab type would you like to see more of?
Hands-on investigative (1)
Computer simulation tutorial (2)
Reading literature (3)
Completing worksheets (4)
None or other (5) ____________________

Q31 How much has the lab has increased your desire to continue learning about this material?
A great deal (1)
A lot (2)
A moderate amount (3)
A little (4)
None at all (5)
Q15 Which specific labs would you recommend retaining for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Motor unit demo (2)
- EMG with weights (3)
- M-wave, H-reflex demo (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Postural control feedforward (10)
- Reflexes & stretching: tendon tap, balls, partnered techniques (11)
- Kinesthesia vibrators (12)
- Locomotion (13)

Q16 Which specific labs would you recommend removing for future semesters? Select all that apply.
- Partner-evoked contractions w/ electrical stimulation (1)
- Motor unit demo (2)
- EMG with weights (3)
- M-wave, H-reflex demo (4)
- Adaptation w/ goggles, balls, pennies, & baskets (5)
- Action potential tutorial (6)
- GHK, Nernst Equation application (7)
- Plasticity & rehabilitation review document (8)
- Neuroprosthesis & grasping review document (9)
- Postural control feedforward (10)
- Reflexes & stretching: tendon tap, balls, partnered techniques (11)
- Kinesthesia vibrators (12)
- Locomotion (13)

Q17 How much do you think the experiences acquired in the lab will impact your future career or educational goals?
A great deal (1)
A lot (2)
A moderate amount (3)
A little (4)
Not at all (5)
Q18 What are your future goals (career/education)?
- MD/DO (1)
- PT (2)
- Nursing (3)
- Education (4)
- Graduate school (5)
- Other (please specify) (6) ____________________

Q19 Why did you enroll in this course (i.e. required course, interested in topic, fit schedule best)? Select all that apply.
- Required for degree (1)
- General interest (2)
- Prepare for future goals (3)
- Fit into schedule best (4)
- Other (please specify) (5) ____________________

Q20 Please provide any additional feedback on the lab portion of this course.
Chapter Nine

Does the extent of active learning present in an IPHY lab course affect the information retention and learning gain of statistical concepts among undergraduate students?

Chris Stamper¹ and Adam Blanford²
¹Dept. of Integrative Physiology; ²Graduate Teacher Program

Introduction

This paper discusses the use of the Differentiated Overt Learning Activities (DOLA) framework (Chi 2009; Menekse et al. 2013) to study a) the extent of “active learning” (with a lowercase “a”) taking place within an Integrative Physiology classroom; and, b) the effects of active learning on the retention and application of statistical information learned in lower-level IPHY classes. “Active learning” is defined as students taking part in their own education through activities designed to strengthen their knowledge, exercise their capabilities, and improve metacognitive skills (Prince 2004; Tanner 2012). This is a contrast to “passive learning,” in which students act as receivers of information, usually through a lecture-based format (Chi 2009; Menekse et al. 2013). Passive learning, although a traditional learning format in the United States, does not necessarily exercise student learning and engage metacognition (Menekse et al. 2013).

The IPHY course, called “Physiology Laboratory,” that was studied was recently redesigned to include a greater emphasis on statistical knowledge and reasoning in laboratory modules. Students taking this course have to exercise their knowledge and apply statistical concepts, in addition to figuring out related concepts such as modifying codes in R and the use of Excel. The differentiations within the DOLA framework provide a clear set of criteria for evaluating how much active learning is taking place in the classroom.
To facilitate this analysis, the authors used the Differentiated Overt Learning Activities (DOLA) framework to analyze the extent of active learning associated with the retention of statistical concepts taking place in a Physiology Laboratory. The DOLA framework consists of four categories with corresponding activities that vary with the extent of active learning taking place (see below).

In this paper, we discuss the results of our research into the extent of active learning and statistical retention within an IPHY laboratory course. We begin with a discussion of the theoretical framework underpinning our study, particularly the DOLA framework and its impact on our observations. Next, we discuss the methods used to analyze active learning and retention, and conclude by presenting our results and future directions. Our analysis shows that there is a considerable amount of active learning taking place in this particular laboratory course, making it more conducive for the recall and application of statistical procedures.

**Theoretical Background**

This study uses the DOLA framework to assess the extent of active learning and retention of statistical knowledge because it provides a series of observable classroom behaviors that differentiate between types of activities and levels of engagement (Menekse et al. 2013). While studies suggest that “active” learning is preferable, they disagree on what constitutes active learning and the overall impact it has on student success. In this section, we discuss the literature on active learning, its limitations, and how the DOLA framework clarifies classroom learning and observational analysis.

Studies show that engaging students in classroom activities activates their conceptual knowledge of a topic while also causing them to reevaluate the ways in which they think about
that topic, leading to a deeper understanding of the material and its applications (Tanner 2012). In contrast, the traditional classroom involves teachers conducting lectures while the students sit and take notes, often with little interaction (Menekse et al. 2013). Students may learn the information for a limited period of time but it is not necessarily committed to long-term memory. As a result, scholars suggest that active learning is the preferred method of instruction.

However, the literature also shows that while active learning is preferable, the term is so broadly defined that it cannot be applied meaningfully to classroom research. For example, one definition of “active learning” is that it involves the students “doing something” (Prince 2004). Other definitions involve either more or less discriminating forms of activity.

The DOLA framework provides a clearer set of criteria (Chi 2009; Menekse et al. 2013). It consists of four categories with corresponding activities that vary with the extent of active learning taking place:

- **Passive**: Students are passive in the learning process, sitting and listening to the teacher or being walked through the lesson step by step.
- **Active** (with uppercase A): Students doing something, such as underlining notes or clarifying directions.
- **Constructive**: Students working together to ask questions, create analogies, or improve conceptual understanding.
- **Interactive**: The most dynamic form of learning, students exhibit a high degree of interaction and participate in interactive activities as well as the co-construction of new knowledge.

The IPHY course that was studied was recently redesigned to include a greater emphasis on statistical knowledge and reasoning in laboratory modules to build on the statistical knowledge
students gained in their introductory-level IPHY statistics courses. Earlier versions of the course tended to rely on a more passive mode of instruction, leading to less meaningful student learning experiences. Students now taking this course have to exercise their knowledge and apply statistical concepts, in addition to figuring out related concepts such as modifying codes in R and the use of Excel to manipulate data and create graphs. The differentiations within the DOLA framework provide a clear set of criteria for evaluating how much active learning is taking place in the classroom.

As the descriptions suggest, the DOLA framework classifies learning based on the type of activity and degree of complexity involved, ranging from simply listening to a lecture (passive) to developing new types of knowledge with fellow students (interactive). This framework clarified our understanding of student behaviors and made it possible to conduct classroom observations, as described in the next section.

Methods

After obtaining informed consent, authors used data from classroom observations, pre/post tests, and student grades to determine the extent of active learning and the retention of student statistical knowledge. In this section, we discuss each of these methods in detail and how they contribute to the study.

Classroom Observations

The authors conducted a total of four observation sessions using a DOLA observation worksheet, two for a morning physiology laboratory session and two for an afternoon session. Given that each lab was four hours long and each session included quizzes, discussion, and data collection,
the observations focused exclusively on the statistical analysis portion of each lab. Each lab allowed a maximum of sixteen students, with most broken up into working groups of four. During the analysis portion, the students worked collectively to analyze data from all of the lab groups using their knowledge of statistics as well as a working knowledge of the R statistical program.

Each lab group was given a letter designation of A, B, C, or D, and the authors observed each group together at five-minute intervals. The authors conducted several five-minute observation periods for each group during the first lab session to ascertain whether they were classifying the same or markedly different behaviors. Accuracy was confirmed using interrater reliability tests immediately after the first day’s observations concluded (see results). Once the accuracy was confirmed for the first lab, the authors conducted a shorter observation period (5 minutes for each group) for the second lab.

Pre- and Post-Test Data

The authors also used pre- and post- assessment data to determine student learning gains in knowledge and retention. The Integrative Physiology department administers a standard twenty-question learning assessment at the beginning and end of each semester to establish a baseline for each class and chart their learning gains over the course of the semester. Students are presented with problems and must choose the appropriate statistical tool to analyze the given data set. The results of the pre-/post- assessments are discussed below.
Final Grades

To gain a comprehensive view of student learning, the authors also used anonymized final grades to determine student performance at the conclusion of the semester.

Analysis

Classroom Observations – Lab One

As stated above, the authors conducted double the amount of observations for the first lab to ascertain interrater reliability, leading to higher numbers for each. The following table shows the data for the morning laboratory groups during the first set of observations:

<table>
<thead>
<tr>
<th></th>
<th>Morning Group A</th>
<th>Morning Group B</th>
<th>Morning Group C</th>
<th>Morning Group D</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Constructive</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>9.6</td>
</tr>
<tr>
<td>Active</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3.9</td>
</tr>
<tr>
<td>Passive</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Table 1: The results of the first set of morning laboratory observations using the DOLA framework.

The table illustrates that the majority of the observed student behaviors were classified as “interactive” or “constructive.” The average number of interactive behaviors was thirteen across all four groups, and largely consisted of students working collectively as well as debating amongst each other about the proper statistical procedures to use in the analysis. The numbers were relatively consistent across all four groups.

The average number of constructive behaviors was 9.6. Constructive behaviors consist of posing research questions, creating charts or graphs to describe the research problem, or comparing the current problem with prior, parallel experiences (Menekse et al 2013). Of the
four morning groups, Group D had nearly twice the number of observable constructive behaviors, as all four group members discussed the statistical procedures and how to translate the analysis into R to achieve results. While the other groups did engage in constructive behaviors, they did not interact to the same extent.

Active and passive behaviors occurred in much lower numbers. The average number of active behaviors was 3.9, and these consisted of consulting the laboratory manual and confirming directions. Group B had three instances of active behaviors, Group 2 had two, and Group A had one. Group D had no instances of active behaviors as they seemed fairly confident in their understanding of the material.

Passive behaviors included more involvement from the instructor and/or learning assistant and the authors saw an average of 2.1 interactions per group. Group A had the most with five passive behaviors, mostly found in the instructor offering unsolicited guidance or walking them through the analysis. Group D had four instances of instructor involvement, while Group B had two. Group C had no instances of passive learning.

These data are similar to those found during the afternoon laboratory session. The following table shows the results of the afternoon observations:

<table>
<thead>
<tr>
<th></th>
<th>Afternoon Group A</th>
<th>Afternoon Group B</th>
<th>Afternoon Group C</th>
<th>Afternoon Group D</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>13</td>
<td>10.5</td>
</tr>
<tr>
<td>Constructive</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>14</td>
<td>9.25</td>
</tr>
<tr>
<td>Active</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Passive</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>3.75</td>
</tr>
</tbody>
</table>

*Table 2: The results of the first set of afternoon laboratory observations using the DOLA framework.*
Like the morning laboratory groups, interactive learning behaviors were the most prevalent with 10.5 behaviors on average. Groups B and D exhibited the same types of behaviors with thirteen instances, working collectively to solve the problem and debating amongst themselves to determine the correct statistical procedures. Group A had ten interactive behaviors, while Group C had six. Group C took longer to reach the analysis stage than other groups, which might explain the lower numbers of observations.

The average number of constructive behaviors was 9.25, with Group D exhibiting the most instances of these with fourteen. Like morning group D, the afternoon group posed research questions and offered examples to stimulate discussion. Similarly, Group B had twelve behaviors, which is interesting given that they had one group member who was out sick that day. Group C had eight constructive behaviors, and Group A had the smallest number with three.

Unlike the morning group, there were more passive behaviors as the instructor and the learning assistant assumed a more prominent role in the instruction. Group D had the most with five behaviors, while Groups A and B each had four. Group C had the least number, perhaps because they were the last to begin the analytical phase.

Classroom Observations - Lab Two

The following table shows the results of the second set of observations for the morning laboratory session. The authors conducted a much shorter observation period given that they had established interrater reliability during the first set of observations.
Table 3: The results of the second set of morning laboratory observations using the DOLA framework.

During the second lab, constructive behaviors were most prevalent with an average of 6.75 behaviors. Group D again exhibited more than twice the number of constructive learning behaviors (12) as they debated the proper statistical procedures and coding necessary to analyze the data. The other three groups – A, B, and C – each had five constructive learning behaviors.

Interactive behaviors were also prevalent with an average of 5.25 behaviors during the observational period. Group D also exhibited twice the number of behaviors for this category as well, as they engaged in debate, collective work, and even reciprocal teaching (i.e. students teaching concepts to each other). Groups A, B, and C exhibited much lower numbers.

Active behaviors averaged 3.25 during the observation period, with Groups A and B exhibiting the most behaviors like referencing the lab manual and confirming the directions for the lab. Group D had one behavior, while Group C had none.

There were few passive behaviors observed during the period, with only Group A and Group B exhibiting two each. Groups C and D showed none of these behaviors.

The last table shows the data for the second observation of the afternoon laboratory groups:
Table 4: The results of the second set of afternoon laboratory observations using the DOLA framework.

The table shows that constructive behaviors were most prevalent with an average of 10.25 behaviors per observation period. Group B had the most constructive behaviors as they posed research questions and discussed different approaches to the statistical analysis. Group A exhibited twelve behaviors in this category as they also engaged in intensive discussion. Groups C and D each had eight instances of constructive behaviors.

The interactive behaviors were the second-most prevalent, with an average of 6.75 behaviors. Group B had ten interactions as they engaged in debate, received interactive feedback from the instructor, and engaged in reciprocal teaching. Groups C and D each had six instances of interactive behavior, while Group A had five.

The active and passive behaviors were very low for this period, with only 1.75 passive behaviors and 0.75 average active behaviors. The instructor had less of a presence during this lab.

In all, these data show that the majority of learning behaviors that students exhibited fell into the “interactive” and “constructive” categories of the DOLA framework, which suggests that the lab’s revised structure is effective in encouraging students to activate their existing knowledge base and solve the problems using their statistics experience. The instances of the instructor and/or learning assistant engaging the students was very low, and perhaps unavoidable since s/he must be available for questions when asked.
Pre- and Post-Test Data

The pre-test was taken at the beginning of the spring semester during the first day of class. The table summarizes the aggregate findings for all 29 students involved in the study:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>45%</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12%</td>
</tr>
<tr>
<td>Lowest Score</td>
<td>20%</td>
</tr>
<tr>
<td>Highest Score</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 5: Pre-test statistics for both the morning and afternoon laboratory sessions.

The mean score for the pre-test was forty-five percent with a standard deviation of approximately twelve percent. The lowest score on the pre-test was twenty percent and the highest was achieved by a single individual with seventy-five percent. These data show that the students began the course with roughly the same knowledge base. The next table shows the results of the post-test:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>71%</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12%</td>
</tr>
<tr>
<td>Lowest Score</td>
<td>45%</td>
</tr>
<tr>
<td>Highest Score</td>
<td>90%</td>
</tr>
</tbody>
</table>

Table 6: Post-test statistics for both the morning and afternoon laboratory sessions.

The table shows a mean of seventy-one percent, with the same twelve-percent standard deviation. The lowest post-test score was forty-five percent and the highest (achieved by the same individual who received the highest pre-test score) was ninety percent. In all students achieved a mean learning gain of twenty-six percent.
Final Grades

The mean final grade was eighty-six percent, with a standard deviation of nearly six percent. The high average supports the premise that the class structure helped to enhance the student learning experience.

Analysis

Each data set shows that there is a high degree of “active” learning taking place in the IPHY physiology laboratory course. For each set of classroom observations, the authors found that student learning behaviors were geared toward the higher levels of the DOLA framework; namely, the constructive and interactive levels. Students felt comfortable discussing the statistical concepts and relevant data with each other, even going so far as to correct misperceptions and explain their reasoning. The behaviors identified as “passive” and “active” were generally low, suggesting that the students felt comfortable working with the data, the associated statistical concepts, and the use of the R statistical program. Moreover, the instructor’s and learning assistants’ involvement was minimal, often restricted to giving advice on how to approach a problem.

In addition, the pre- and post-test data show a marked improvement in student learning gains from the beginning to the end of the course. Given that the assessment was designed to ascertain students’ disciplinary and statistical knowledge levels, the authors suggest that the students were able to take the knowledge they gained in the lower-level statistics courses, internalize it, and apply it productively to the solution of the novel problems presented by each lab.

The final grades show that the students grasped the concepts and were able to apply them productively in the classroom. The grades are notable for the high average score of eighty-six percent and the fact that the majority of the students completed the course with A’s and B’s.
One interesting trend we discovered in analyzing all of these data together was the existence of a stable relationship between students who achieved scores in the upper 35% and the lower 35% of the class grade distributions. Performance on the pre-test was often indicative of where they would end up at the conclusion of the course, as students who achieved high scores on the pre-test finished with high A’s, and those who did not achieve a high score on the pre-test finished with low B’s or C’s. The following graph illustrates the relationship:

![Figure 1: Bar graph illustrating the stable relationship between students in the upper and lower 35% of the class score distribution.](image)

The bar graph shows the distribution of students’ grades in the upper and lower 35% of the course. Students who achieved high scores in the upper 35% of the pre-test distribution continued to achieve higher scores and exhibited greater improvement overall than students who scored in the lower 35%. One possible direction for future research would be to ascertain if the
students in the upper 35% of the course also exhibited higher degrees of interaction with the active learning environment.

**Conclusions**

In this project, the authors used the Differentiated Overt Learning Activities (DOLA) framework (Chi 2009; Menekse et al. 2013) to study a) the extent of “active learning” (with a lowercase “a”) taking place within an Integrative Physiology classroom; and, b) the effects of active learning on the retention and application of statistical information learned in lower-level IPHY classes. Our investigations showed that a high degree of active learning was taking place and students were able to apply statistical concepts they learned in lower-level courses to the solution of novel problems in integrative physiology.

Future investigations should focus on increasing the sample size and the length of the study to encompass additional laboratory sessions. Looking at more classes and over a broader period of time will help determine if the high degree of active learning is prevalent throughout the semester. In addition, we recommend expanding the observation period to encompass at least part of the data collection portion of the lab, as students may discuss statistical concepts in preparation for their later data analysis.

**Impact on researcher**

Prior to this project, I had very little experience with educational research. I found this experience to be exceptionally rewarding and an excellent opportunity to advance and hone my capabilities as a well-rounded researcher. My graduate work mainly focuses on bacteria in the gut and how they interact with the brain. Needless to say, this project immersed me into a totally
unfamiliar sector of research. I had never performed in class observations before, which was one of the most rewarding aspects of the project. I experienced the classroom in a whole new light and was able to critically examine different ways and styles of learning. Reflecting on these in-class observations has allowed me to reform my teaching to cater to different learning styles more so than attending any seminar or talk on pedagogy. One relatively unexpected benefit was the experience of being a PI on a project that I developed and carried out. I really got an insider’s perspective of all aspects of the experimental process including outlining the project, writing IRB protocols, developing an observation score sheet, running the experiment, analyzing the results, and writing up the project. I know that all of these exercises have and will continue to assist in my research pursuits.

**References Cited**


Chapter Ten

Characterizing the Accessibility Climate in Geology

Emily Fairfax and Megan R.M. Brown
Department of Geological Sciences, University of Colorado Boulder

Introduction

The goal of this study is to quantify the “accessibility climate” of the Department of Geological Sciences at University of Colorado Boulder. We define accessibility climate as the combination of a person’s experiences and perceptions of issues rooted in accessibility, the ability for students and/or faculty to participate regardless of disability or limitations, within or in relation to the department. We have observed that few students with physical limitations enroll in geology labs. In addition, students not able to participate in field trips are often excused without accommodations that make the lab more inclusive and give the disabled student an equitable experience.

Currently, there are no quantitative data describing how the Department of Geological Sciences at CU Boulder handles issues of accessibility and disability. There are also no data describing the experiences of undergraduates enrolled in our geology courses specifically as it relates to accessibility. Through surveys, we characterize the accessibility climate within the department and determine in what ways the current accessibility climate is related to our enrollment statistics for students with physical disabilities. This characterization is useful for teaching personnel awareness of accessibility issues and efforts to be more inclusive as they go forward in their teaching. It also fosters trust between the undergraduate students and the teaching personnel in the Department of Geological Sciences at CU Boulder. The survey and changes made based on the results show that the teaching personnel are making a sincere effort
to listen to the experiences and perceptions of undergraduates within the department regarding issues of accessibility.

**Prior Experience with Education Research**

Although we are comfortable with and experienced in geoscience research, neither of us had completed education research or a Teaching as Research (TAR) project before. We have a long-standing interest in discipline-based education research (DBER) and took multiple elective courses in psychology, sociology, and other social sciences during our undergraduate degrees at Carleton College (Emily) and Arizona State University (Megan). Additionally, we spent several months thinking about the questions and problems we wanted to investigate. We also pursued more information by attending talks, posters, and workshops on related content at the annual meetings for the American Geophysical Union and the Geological Society of America in fall 2016. When we realized that the project we had in mind was similar in content to many of the DBER projects presented at the conferences, we decided to formalize our research through the TAR program.

**Our Expectations Prior to Beginning the TAR Project**

We expected that doing formal education research through a TAR project would be challenging simply because it is something we were not entirely familiar with. We knew going into the project that we would need to brush up on statistics, read more background papers on geoscience education research, and learn the logistics of how to survey students and do research that involves human subjects.

We also expected that we would have ample support for our project through both the Graduate Teacher Program (GTP) office and our home department. We attended numerous GTP
workshops and events prior to committing to a TAR project. We knew the level of involvement and availability that various personnel within the GTP would have. Our first meeting with our research advisor, Dr. Adam Blanford, confirmed that the TAR program and GTP in general would guide us through learning the ropes of education research and help us develop and refine our big ideas into actionable research questions.

Theoretical Background

National statistics illustrate the number of individuals with disabilities studying or working in STEM disciplines is dwindling (NSF 2012). For example, Atchison and Libarkin (2013) report that only seven percent of STEM practitioners and one-half of one percent of college students report having a disability. Among the explanations proposed for such a small number is the impact of negative prior experiences with teachers and instructors who were unable or unwilling to accommodate a disability, making it difficult for students with disabilities to fully participate in classroom activities.

Research suggests that correcting this situation must start at the pedagogical level - retraining TAs and faculty to design inclusive curriculum and implement equitable accommodations for students with physical disabilities, and redesigning existing curriculum to be more in accordance with the principles of Universal Design (Shaw 2011; Atchinson and Libarkin 2013). Inclusive and Universal Design are simply guidelines that have been established to help educators design curriculum, spaces, and the learning environment to provide an equitable experience for students from the start. This is in contrast to relying on numerous accommodations that may be requested with little or no advanced notice. Some examples of Inclusive and Universal Design include making sure all notes and PowerPoints can be read
through a screen reading device for the visually impaired; making sure that there are wheelchair accessible desks in all classroom spaces; and using transportation for field trips that is accessible for students with all levels of mobility. An important element of inclusive accessibility is that if an unexpected issue arises, educators and students need to have an open and honest conversation about what accommodations are needed and how they will be provided.

If the faculty and TAs are more conscious of issues in accessibility when designing curriculum, they will design better labs and classes from an inclusion standpoint. Students with physical disabilities will feel more comfortable and perceive the curriculum more positively, which bolsters their learning and overall experience (Cook et al. 1997). This could potentially increase enrollment and retention of students with physical disabilities. More visibility of students with physical disabilities in the department will further increase awareness of accessibility issues, which the now better trained faculty and TA’s will be equipped to address.

Using the results of our surveys, we can develop action plans for increasing enrollment of undergraduate students with physical disabilities in the Department of Geological Sciences at CU Boulder. For example, the positive feedback loop between the two populations - students and teaching personnel - can be used to our advantage. If either the number of students with physical disabilities increases or if the teaching personnel receive better training, we expect a more positive accessibility climate to develop. That improved climate goes on to attract more students and/or increase the demand for better training. Additionally, if a common experience or perception of the Department of Geological Sciences is that existing curriculum is inaccessible, then the curriculum can be systematically reviewed and updated to be in accordance with the principles of Inclusive and Universal Design.
Methods

Survey Creation

We took great care to write a survey that had unbiased questions, did not have the potential to cause significant emotional harm, and would not collect any personally identifying information. Some examples of the questions are as follows:

- (Teaching Personnel) *Rate your agreement with the following statement: The labs I teach at CU are accessible to students with physical disabilities.*

- (Undergraduates) *In your opinion, are field trips accessible to a student with a disability?*

We constructed the survey using the Qualtrics survey platform, with questions written by the authors and reviewed by our research advisor, Dr. Blanford. The questions went through multiple revisions before being finalized. We tested the survey for accessibility and technical issues before deploying the survey. Megan took charge of her specialized area of interest within the project - undergraduate experiences. Emily took control of her specialized area of interest - knowledge and training of the teaching personnel.

Within Qualtrics, we built survey logic so that participants were not shown questions that were not relevant to them or their experiences. For example, anyone who identified as an undergraduate was taken to a block of the survey that focused on personal experiences in the department with disability and accessibility. Anyone who did not identify as an undergraduate was taken to a block of the survey that focused on teaching students with physical disabilities in class and lab settings. The participants taking the undergraduate portion of the survey never had the option to answer the questions designed for teaching personnel.
Additionally, when designing the survey we intentionally collected purely quantitative data (i.e. “rate your agreement with this statement on a scale of 1-10.”; “how many years have you been teaching.”; etc.) as well as qualitative free-response data. When thinking about the pros and cons of each type of data, we ultimately decided that the best course of action was to collect both types to help us build a more complete picture of the accessibility climate in our department.

**Subject Recruitment and Survey Deployment**

Subjects for our research were drawn from the lists of undergraduate students enrolled in introductory geology classes (GEOL 1010 and GEOL 1030), undergraduate students who are geology majors, and teaching personnel (including faculty, instructors, and teaching assistants) within the Department of Geological Sciences.

For those students enrolled in introductory geology courses, we identified all instructors for the classes being taught during the current semester and contacted them about coming to their classes for 5-10 minutes to explain the research. After explaining the research, we had the instructor of record send the survey and a pre-written, Institutional Review Board approved, recruitment email to all of the undergraduates in the class. We also had the department administrative staff send the recruitment email and survey to all students registered as geology majors via a department listserv.

In order to garner sufficient faculty response rates in a relatively small department, we attended a faculty meeting to explain the research and request participation. After explaining the project, the survey was sent out to the faculty listserv by the administrative staff in the department. We also sent the recruitment email and survey to all graduate student TAs, postdocs,
and instructors who have taught an undergraduate geology class in the last 5 years via the administrative staff. We understood that surveys are getting sent out over listservs nearly every day and that people are busy, so we sent reminder emails to all participants at two weeks prior to survey closing and again at one week prior to survey closing.

We took care to follow guidelines set forth by the Institutional Review Board for research on human subjects throughout the process. For example, the subjects were not under the instruction/supervision of either of us; we made it clear that participation would not affect grades for students or performance evaluations for teaching personnel; no one under 18 was permitted to take the survey; and everyone had the option to drop out of the study at any time.

**Data Collection and Analysis**

Data was collected through Qualtrics, the same software that was used in the creation of the survey. All data was collected anonymously to prevent any issues with personally identifying information. It was then exported in its raw form onto a password protected drive to ensure a pristine data set remains available throughout the data analysis and coding of results.

We coded our free response answers to identify common themes, ideas, or feelings that people were expressing in response to the prompts (e.g. students uncomfortable approaching instructor, confusion about accommodation responsibilities, and the department as a whole has room to improve accessibility). To ensure inter-coder reliability, Emily coded the teaching personnel’s responses and Megan coded the undergraduate responses and we blind-coded each other’s data sets.

For the quantitative data, much of it was analyzed as bar plots representing the frequency of each response. In the teaching personnel portion of the data, the responses were analyzed as a
whole, and then also according to role in the department - i.e. TA, faculty, instructor, etc. Emily was interested to see whether everyone who teaches has similar experiences and knowledge of issues in accessibility, or if there is a divide between subgroups. Similarly, Megan analyzed the undergraduate portion of the data as a whole and then according to other factors, for example, individuals identifying as having a disability and/or the academic major of the subjects.

Results

The results of our study indicate that amongst teaching personnel there is a general lack of knowledge of the principles of Inclusive and Universal Design, as well as a lack of training and experience in actually accommodating or designing curriculum for students who have physical disabilities. Teaching personnel survey respondents expressed that although they recognized that there were issues in the course design and curriculum when it comes to accessibility and equity, they did not feel adequately prepared to handle the issues themselves, or even really knew what was expected of them legally or by the department. Student respondents indicated that the accessibility climate had room for improvement. Most students surveyed found field trips less accessible than the classroom with the physical terrain and transportation cited as main issues.

Participation is Important, but Not Always Easy

Of the 48 teaching personnel respondents, 60% said that they would typically make participation a part of a student’s grade in their lectures, and 80% said participation would be a component of the student’s grade in a lab course. In an open response question that was coded, 67% of respondents indicated that attendance counted toward the participation grade. However, only 6 of the 48 participants reported that they had any knowledge about what Universal or Inclusive
Design is. This may indicate that few courses are being designed to be accessible for students of all ability levels from the start. For students with physical disabilities, simply attending classes with field trips or labs that go outside and were not intentionally designed to be as accessible as possible can be a significant challenge.

This sentiment was echoed in the undergraduate portion of the survey, where when asked about the accessibility of lab courses the majority of the participants talked about the difficulty in the terrain, the physical requirements to hike at altitude, the fact that many of the busses that transport students to field sites are not accessible to wheelchairs and are difficult to use with crutches. In addition, the reasons for not asking for accommodations (including not feeling like the accommodations would be made; the disability not feeling “real” enough; and trying to accomplish the tasks on the same terms as everybody else) are disheartening.

Acknowledgement of the Issue, Lack of Knowledge to Fix It

The majority of participants (73%) in the teaching personnel portion of the survey acknowledged that the Department of Geological Sciences at CU Boulder is not enrolling the number of students with physical disabilities that would be expected based on the demographics of the undergraduate population as a whole. Additionally, nineteen of the forty-eight (40%) teaching personnel respondents indicated that they did not think the lab courses they taught were accessible, while thirteen of forty-eight (27%) did think they were accessible. When broken down by sub-group within the teaching personnel portion of the survey, twelve out of the original nineteen (63%) respondents who did not think their labs were accessible were teaching assistants. In the Department of Geological Sciences, most TAs are the instructor for the introductory level labs, so the fact that more TAs are finding their labs inaccessible than faculty
may be indicative of more knowledge of accessibility issues, or it may be that the intro labs are particularly inaccessible in nature. In the undergraduate experience portion of the survey, forty of the seventy-six (53%) undergraduate respondents indicate that field trips are inaccessible. Of the fifty-eight undergraduate respondents who said they had taken a class with a field trip, approximately sixty-six percent said in their opinion the field trips were not accessible for students with a disability.

**Conclusions from the Research**

Although there is still work to be done with the large dataset we collected through our TAR project, the preliminary results have made it clear that a good first step for our department is to reevaluate how we are training the faculty and TAs on accessibility issues. There will need to be discussions between us, the faculty, the TAs, and the department chair about what route forward we think will have the best chance of success. Possible options include adding new trainings to the fall TA training; having more frequent short workshops/brush-up trainings for all current teaching personnel; and redesigning some of the more problematic courses or field trips identified by undergraduates and teaching personnel as being particular inaccessible to be more in accordance with the principles of Inclusive and Universal Design.

We are happy to see that the general consensus from the survey results is that instructors are not intentionally making classes inaccessible and that faculty/TAs are largely interested in being instructed on what to do and how to increase accessibility. We think that the openness of the department to change will be incredibly beneficial as we move forward to make the geology program - and in particular the introductory geology courses with field trips - more accessible in the future.
Summary of Experience Doing a TAR Project

Having the opportunity to engage in formal education research was challenging and rewarding. We had our share of struggles along the way with revising the Institutional Review Board protocols over and over again prior to submission; learning to use Qualtrics to collect data and design unbiased survey questions; and figuring out how to actually analyze qualitative free response data. With each struggle came a unique reward. After the Institutional Review Board protocols were submitted and approved, we felt much more confident in ourselves and our methods. After learning to use Qualtrics and write survey questions, we quickly were able to put together a complex survey with a variety of question types and complicated question logic. After analyzing the quantitative and qualitative data, we have gained incredibly useful into not only what people’s experiences and knowledge about accessibility issues is, but also why they feel the way they do. Putting numbers on human experiences is no easy task, and it is easy to misrepresent or extrapolate too far when interpreting free-response style questions. One of the things that was new to us and wound up being immensely helpful was coding the free-response questions to look for common themes, sentiments, and ideas that were being expressed by various individuals in their own words.

Overall, we found doing education research through the TAR program to be an overwhelmingly positive experience that has influenced the way we think about research as a whole. We came from very quantitative backgrounds where every statement made needs to be backed up by data. By collecting rigorous data on the topic, we feel that we are now better equipped to propose solutions to the various problems that surfaced in the responses to our survey.
**Future Direction for our Education Research**

Many avenues for further work stem from the data we have collected already. We are looking to implementing various changes to the training TAs and faculty receive on accessibility and disability. A good follow up study would evaluate the efficacy of the changes by surveying the teaching personnel again after the updated training implementation to see if their perceptions, knowledge, or ideas about accessibility in geology had changed at all.

Another option is to rework a number of the more inaccessible labs to be more equitable and in accordance with the principles of Inclusive and Universal Design. We could compare student learning (via grade records and FCQ feedback) in the redesigned courses to the historical trends. Ideally, redesigning the courses would have a positive impact on student learning across demographics, but also in particular for students with disabilities who had previously responded that they did not find the introductory geology courses to be accessible to them. One possibility we have begun to think about is having an interactive, virtual field experience available for students who cannot participate fully in the field trip labs. We also have considered doing a full review of all course material for the geology classes to ensure that they are screen reader accessible. Again, evaluating how well these changes work and what kind of a difference students and teaching personnel are experiencing would be something we’d like to evaluate through formal education research.

In our opinion, any work that is done to better student learning or make education more accessible for all people regardless of race, sex, gender, disability, age, etc. should be peer-reviewed and published so that other institutions can build on the successes and learn from the shortcomings of the rest of the academic community. Peer-review ensures the data was collected in a responsible and reproducible way, and publishing in academic journals makes the data and
results available to all institutions. We intend to submit the full results of this study for publication. We will also continue to be active in education research by repeating the process we experienced with our TAR project: obtaining Institutional Review Board approval, writing unbiased questions, handling data responsibly, and publishing our results.
References Cited


Chapter Eleven

Effect of Learning-by-Teaching in a Flipped Classroom

Victoria Li¹, Wendy Bailey², & Ray Littlejohn²
¹Department of Applied Math, CU Boulder, USA, ²Lockheed Martin Engineering Management Program, CU Boulder, USA

Motivation

A growing US university student population will require adaptation of teaching styles in order to meet educational demands without sacrificing quality for quantity. According to the National Center for Education Statistics, enrollment in degree-granting institutions between 2002 and 2012 increased 24 percent, from 16.6 million to 20.6 million (U.S. Dept. of Education 2015). Specifically, undergraduate enrollment increased 24 percent and post baccalaureate enrollment increased 24 percent. Given constraints on university resources in terms of time and space, how can we maintain the quality of learning and teaching experiences? Research shows that large teaching loads correlate to decreased teacher satisfaction and likewise large lecture sizes correlate to decreased student satisfaction (Monks and Schmidt 2010). Low student satisfaction correlates to decreased student retention rates (Cuseo 2007). The reasons, simply explained, are that students need to be treated as individuals not as numbers, and education is meant to be an experience not an industry.

With the creation of Khan Academy in 2006 emerged the idea of the flipped classroom as a promising teaching style (Fink 2011). Rather than having students listen to lectures in class and do homework on their own, as in traditional teaching styles, flipped classrooms have students watch lectures on their own and work on problems in class. There are many variations of the flipped classroom, but it typically consists of readings/videos/reading questions before class
Advantages of the flipped classroom include increased student responsibility for learning, higher scores, teamwork, development of interpersonal skills, student innovation, ability to learn at one’s own pace, and ability to handle the increased student numbers and decreased state/institutional funding in today’s education system (O’Flaherty and Phillips 2015). The success of the flipped classroom may be explained by the generation effect and the testing effect, both of which have been found to improve students’ learning, and both of which are included in the flipped classroom (Winquist and Carlson 2014). The generation effect is the finding that knowledge generated by students is recalled at a higher rate than knowledge given to them from reading (Slamecka and Graf 1978). For example, having students generate the antonym to a word (i.e. given “hot”, generate “cold”) makes them remember the word pair better than having them simply read it (i.e. “hot-cold”). Meanwhile the testing effect shows that testing can also be used to improve students’ long-term memory and learning, for example students who answer test questions prior to a final test outperform students who study for the same test (Bertsch et al. 2007). Indeed, a study comparing a traditional version of an introductory undergraduate statistics course to a flipped classroom version that used pre-reading questions (testing) and in-class exercises (generating knowledge) found that flipped classroom students did significantly better on the Psychology Area Concentration Achievement Test (ACAT) Statistics scale a year later, even accounting for student quality (ACAT non-Statistics scale scores).

There are also a number of disadvantages to flipped classrooms (O’Flaherty and Phillips 2015): The main complaint from students, when it did arise, was that there was too much student responsibility for learning. Instructor complaints mainly centered around the enormous amount
of up-front preparation required. Use of technology for in-class activities is not necessarily a
good thing: A flipped classroom based on an automated tutoring system had subsequent issues
with grasp of concepts (Strayer 2012), and a flipped classroom using apps to highlight teaching
points showed no difference in effectiveness from a traditional classroom (Martin et al. 2013).
However, the flipped classroom studies that used clickers or pre-lecture videos were met with
improved grades and student satisfaction (Ferreri and O’Connor 2013; McLaughlin et al. 2013;
Gilboy et al. 2015; Yeung and O’Malley 2014). Previous findings demonstrated the greater
importance of content and student engagement over particular resources used, and showed a
t tendency of online resources to improve lower order rather than higher order cognitive skills.
Several studies suggested a need for IT support personnel to prepare videos and content, and a
need for “professional educators” to address issues of instructor ability and understanding
regarding flipped classrooms. Lastly, while there is much qualitative data on the effectiveness of
flipped classrooms, there are few quantitative studies of their effectiveness, and results tend to be
limited to moderate improvements in exam scores (O’Flaherty and Phillips 2015; Winquist and
Carlson 2014).

**Purpose**

The flipped classroom model implemented in our study could offer a solution to the above
problems that also retains the advantages of flipped classrooms. EMEN 5005, Introduction to
Applied Statistical Methods, is a flipped classroom course in the University of Colorado Boulder
Engineering Management Program that uses pre-reading/questions before class, clicker questions
(testing effect) during class, and homework after class. We proposed adding a learn-by-teaching
activity (generation effect) to help the course take full advantage of what flipped classrooms
have to offer. By providing students with a real-life data set (College Scorecard (U.S. Dept. of Education 2016)) and having them generate and answer their own clicker questions about the lecture material (IU8: One-Sample Continuous and Discrete Hypothesis Testing), we engaged students’ higher cognitive skills in a research-like activity. This could help increase students’ responsibility for their own learning, balanced by instructor guidance and examples that were provided for them. In addition, the use of an online dataset integrated technology at reasonable levels. There was a low amount of preparation required, with no need for professional educator support. Quantitative analysis of results was based on grades (IU8 homework scores) and a control (non-IU8 homework scores) to account for quality of students, while qualitative results were obtained through a student survey. If successful, such a learn-by-teaching method could be considered as a solution to declining student retention amidst a growing student population.

**Methods**

We implemented a learn-by-teaching activity in the fall 2015 EMEN 5005 class taught by Dr. Ray Littlejohn, and used data from the spring 2015 and Fall 2014 EMEN 5005 classes, also taught by Dr. Ray Littlejohn, for control analyses. The fall 2015 class was comprised of 33 distance learner and on campus graduate students, while the spring 2015 and fall 2014 classes were comprised of a mix of distance/on campus and graduate/undergraduate students totaling 63 and 69 students respectively. Of the 63 students in spring 2015, there were 52 graduate students. While all fall 2015 on campus students participated in the activity, among them 15 agreed to participate in the study, resulting in a study sample size of 15.

The above three classes were all taught using a flipped classroom lecture style consisting of pre-readings/questions before class, clicker questions during class, and homework after class.
Only the Fall 2015 IU8 (Instructional Unit 8) lectures received the learn-by-teaching activity:

Students were given a subset (CollegeScorecard2013Data.xlsx, CollegeScorecardDataDictionary.xlsx) of the College Scorecard dataset and given the entire class period (75 minutes) to come up with questions relating the lecture material to the dataset. In addition, example clicker questions were provided (examples.pptx). Students were free to collaborate, discuss, and receive guidance from the professor (Dr. Littlejohn), faculty advisor (Wendy Bailey), and PI (Victoria Li) throughout the activity. Once students had generated enough questions, they tried answering them using MVPStats software and their knowledge of hypothesis testing. Figure 1 shows the timeline of the study.

---

**Figure 1:** Timeline of learn-by-teaching study.

- **Nov 3** - Introduce project. Students sign informed consent forms.
- **Nov 5** - Learn-by-teach activity 1: One-Sample Continuous Hypothesis Testing. IU8 homework 1 assigned.
- **Nov 10** - Learn-by-teach activity 2: One-Sample Discrete Hypothesis Testing
- **Nov 12** - IU8 homework 1 due. IU8 homework 2 assigned.
- **Nov 19** - IU8 homework 2 due. Student survey.
The learn-by-teaching activity was a rewarding experience for both students and teachers. Students were able to learn practical skills in MVPStats and Excel while solving problems, with the professor present to troubleshoot their work (Figure 2).

**Q:** What do you do when you encounter NULL and Private values in the data?
**A:** Set the value to -999 and change MVPStats settings to only include positive data values.

**Q:** What do you do when MVPStats says "Error: sample size must be n>3" but the sample size is >3?
**A:** Use the comparison matrix to calculate correlation/p-value instead. We then verified the result using R and Python.

**Q:** Two groups both calculated the correlation between control of institution and tuition, and got different results. This happened using both in-state and out-of-state tuition.
**A:** This is because the point-biserial test can only be used for two-outcome nominal data, and control of institution is more than two-outcome.

The professor also showed students shortcuts in Excel.

**Figure 2:** Student questions during learn-by-teaching activity.

During the course of the activity, students got a sneak peek into a future lecture on regression from an interesting fact shared by the professor (Figure 3).

Create a scatterplot of the data, convert the x and y values to z-scores, and the slope will be the correlation coefficient (works for continuous data only).

**Figure 3:** Interesting fact shared by professor.

Some student-generated questions and answers are shown in Figure 4.
As the survey results will show, these types of interactions were key reasons for student satisfaction with learning, while prolonged lack of interaction caused student complaints.

**Quantitative Results**

Anonymized homework scores were collected for all three classes: fall 2015, spring 2015, and fall 2014. We determined the IU8 and average non-IU8 homework score for each student, tested the samples for normality, compared average homework scores across the three classes (Figure 5), and conducted power analysis. For homework score comparison, we chose not to use ANOVA because: 1) IU8 and non-IU8 homework scores are repeated measures, 2) the sample sizes were unequal (different numbers of students in each class), and 3) due to low sample sizes we could not equalize the sample sizes by throwing out data. The techniques required to handle this type of ANOVA would have complicated the ANOVA analysis beyond our means. Instead, we used Welch’s t test to compare fall 2015 (study participants) and spring 2015 (graduate students) IU8 scores, Welch’s t test to compare fall 2015 (study participants) and spring 2015 (all students) non-IU8 scores, and z test to compare spring 2015 (all students) and fall 2014 (all students) non-IU8 scores.
Pearson’s chi-square tests for normality resulted in rejection of normality of the spring 2015 (graduate students), spring 2015 (all students), and fall 2014 (all students) non-IU8 homework data (all p < 0.05). In all cases we assumed that the distribution of IU8 homework scores was the same as that of non-IU8 homework scores. Since the sample sizes for the above samples were 52, 63, and 69 respectively, however, we assumed the samples were normal by the Central Limit Theorem. With a sample size of 15, the fall 2015 (study participants) data could not be tested for normality (df = 0). However, it is reasonable to assume that student grades have a normal distribution, and we made that assumption here.

Although fall 2015 (study participants) IU8 homework scores ($\bar{x} = 96.3, s = 4.7$) were higher on average than spring 2015 (graduate students) IU8 homework scores ($\bar{x} = 95.9, s = 7.3$), Welch’s t test shows that they were not significantly different (p < 0.05). To account for differences in student quality, we compared fall 2015 (study participants) non-IU8 homework scores ($\bar{x} = 90.0, s = 4.2$) to spring 2015 (all students) non-IU8 homework scores ($\bar{x} = 88.7, s = 7.3$), and verified by Welch’s t test that they were not significantly different (p < 0.05). Lastly, we were interested in whether fall and spring semester grades differed historically. Comparison of spring 2015 (all students) non-IU8 homework scores ($\bar{x} = 88.7, s = 7.3$) to fall 2014 (all students) non-IU8 homework scores ($\bar{x} = 84.1, s = 12.8$) by z test found that spring 2015 grades were significantly higher than fall 2014 grades (p < 0.05). Results are summarized in Figure 5.
The power of the test was computed for each of the above three statistical tests (Figure 6). The fall 2015 vs. spring 2015 IU8 comparison had effect size 0.063, power 0.055, and significance level 0.05. The low effect size and power was due to the small study sample size of 15 and a difference in sample mean homework scores of only .4 points, whereas the study was designed to detect homework differences of 5 points (which would yield effect size 7.3 and power 0.69 in this case). Thus there was a high probability of incorrectly failing to reject the null hypothesis. Similarly, the fall 2015 vs. spring 2015 non-IU8 comparison had effect size 0.17 and

*Figure 5: Statistical test results.*
power only 0.091. The spring 2015 vs. fall 2014 non-IU8 comparison, however, had large sample sizes 63 and 69, effect size 0.44, and power 0.81. This met the standard for acceptable power of a test.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Test</th>
<th>Effect Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2015 vs. Spring 2015 IU8</td>
<td>Welch's t</td>
<td>0.063</td>
<td>0.055</td>
</tr>
<tr>
<td>Fall 2015 vs. Spring 2015 non-IU8</td>
<td>Welch's t</td>
<td>0.17</td>
<td>0.091</td>
</tr>
<tr>
<td>Spring 2015 vs. Fall 2014 non-IU8</td>
<td>z</td>
<td>0.44</td>
<td>0.81</td>
</tr>
</tbody>
</table>

*Figure 6: Power analysis results.*

**Survey Results**

Study participants were surveyed to qualitatively understand their attitudes about the learn-by-teaching experience. All study participants filled out the survey, and their responses are attached (survey responses.pdf). Overall, the result was that 9 of 15 students (60%) would prefer their classes be taught with the learn-by-teaching lecture style, if they had to choose yes or no. Whatever their preference, though, almost all students agreed that the time given was more than what was necessary for the activity. The other complaint from students was the lack of structure. Suggestions for improving this were:

- Go over the example questions and their solutions first (we chose to display them but not go over them).
- Make the activity more challenging.
- Make use of the product (for example have students present their work in front of the class).
- Conclude with a summary and discussion.
Interestingly, the amount of interaction with instructors and peers seemed to be the major determinant of student satisfaction/dissatisfaction. Students enjoyed working in a stress-free and flexible environment where they could practice on their own, discuss with peers, and have quick access to instructor guidance when needed. Students did not enjoy lack of feedback and not knowing whether they answered the question right or not. However, only two students encountered this situation, indicating that the majority of students benefitted from increased interaction in the learning activity. Regarding learning outcomes, students generally wrote that the activity taught them to 1) apply MVPStats and 2) make and test statistical hypotheses. In addition, three students wrote that they learned to use the appropriate correlation test for the appropriate data type involved.

**Discussion**

Quantitatively, the learn-by-teaching activity did not significantly change students’ homework scores either for better or for worse. This is probably due to the fact that the class was a flipped classroom to begin with, and the learn-by-teaching activity only made it slightly more so. Qualitatively, the learn-by-teaching activity primarily affected student-student and student-teacher interactions, as well as students’ abilities to apply MVPStats software for statistical testing. The general consensus from survey results was that the learn-by-teaching activity taught hands-on skills, balanced individual work with peer and instructor guidance, and could benefit from more structure. Student survey results indicate that success of the flipped classroom is contingent upon the amount of interaction achieved.

Our findings indicate that learning-by-teaching in flipped classrooms would be suitable for a growing university population. This combination of teaching styles requires low
preparation time, allows students to benefit from interactions with each other and instructors, emphasizes application and troubleshooting of knowledge, and requires less classroom time to achieve the same learning outcomes. By providing students with guidance where they need it most, while decreasing in-class lecture load of instructors, learning-by-teaching could potentially maintain student satisfaction and quality of learning in large classrooms. However, our quantitative study was unable to obtain a large enough study sample size to achieve adequate statistical power, nor to reasonably comment on the effectiveness of this teaching style for large classroom sizes. Future work should also study instructors’ perspectives on this teaching style. Thus future studies in large classrooms and diverse fields will be needed to support our findings.

**TAR Experience**

My TAR experience has been eye opening and very positive from beginning to end, partly because of the people, partly because of the research. In terms of people, Laura Border and Adam Blanford have a contagious enthusiasm for teaching that truly comes from the heart, and really know what they’re doing. From summer workshop to project mentorship to meeting old and new TAR students to post-project follow-up, this has been a well-organized and highly educational year and a half experience. I love how the focus is on research (hence, Teaching As Research) because the best way to learn teaching well is to apply it.

I have always been fascinated by the flipped classroom, but never encountered it at CU until TAR gave me the chance to do research on it. One of the workshop speakers, Wendy Bailey, teaches flipped classrooms at FRCC, and her perspectives on engineering management connected with me on a deep level. She introduced me to Dr. Ray Littlejohn, who kindly agreed to help us conduct a study on EMEN 5005 Intro to Applied Statistical Methods. From planning
to IRB approval to classroom implementation to results, I benefited greatly from their statistics and teaching expertise, and gained experience being a project PI.

I learned many things from my TAR project, including the result of the study itself. They are:

- How to write a protocol
- How to conduct power tests and tests for normality
- When ANOVA should and should not be used
- Application of statistics vs. theory of statistics
- Engineering Management vs. Applied mathematics approaches to statistics
- How to calculate a correlation coefficient from the slope of a line
- Students liked having on-demand guidance and problem-solving in class
- Students requested more instruction at the beginning of the class and a chance to present/make use of their work at the end of the class
- Quantitatively, the learning-by-teaching class style did not significantly improve or worsen student homework scores
- Qualitatively, the effectiveness of the learning-by-teaching class style increased with amount of interaction
- The flipped classroom can benefit students, teachers, and administrators by making education enjoyable and convenient while effectively using class time and space

Looking back, it’s pretty cool to have contributed to cutting edge education research, and to be publishing my results in this TAR Edited Volume.
Special Thanks

We give special thanks to Laura Border, director of CU Boulder CIRTL and Co-PI of CU Boulder TIGER, and to Adam Blanford, TIGER TAR project advisor, for their mentorship.

Funding: TIGER TAR is funded by CU Boulder CIRTL (NSF grant #1231286). The funders had no role in data collection and analysis, decision to publish, or preparation of this manuscript.

References Cited


