CADRE 2018
FEBRUARY 6

THE RELATIONSHIP BETWEEN THE LEARNING ASSISTANT MODEL AND FAILURE IN STEM GATEWAY COURSES

Submitted to: The Learning Assistant Program University of Colorado Boulder

By: Jessica L. Alzen, Laurie Langdon, and Valerie K. Otero
Center for Assessment, Design, Research and Evaluation (CADRE)
School of Education, University of Colorado Boulder

CADRE

ABSTRACT

Course failure rates are an important marker of success for institutions of higher education. Although course failure is not necessarily a rare occurrence across college-level courses, large, lecture-style classes in science, technology, engineering, and mathematics (STEM) departments (i.e. “gateway courses”) often suffer from higher failure rates than other college-level courses. The Learning Assistant (LA) program is designed to facilitate institutional change in ways that help to mitigate course failure rates. Part of this program includes undergraduate “Learning Assistants” (LAs) who support students in their learning in several strategically selected STEM gateway courses. In this study, we examine the relationship between the presence of LAs and course failure rates. We find that students who receive LA support have lower failure rates in gateway courses in the applied math department at the University of Colorado Boulder compared to those who do not receive LA support, but not in the math, chemistry, and physics departments. We hypothesize that the difference in relationships across departments is due to differential implementation of the LA program. We close with a discussion of further work that might provide more insight into the causal mechanism of the LA program on student outcomes such as course failure.

The suggested citation for this report is as follows:

Alzen, J.L., Langdon, L., and Otero, V.K. (2018). The relationship between the Learning Assistant Model and failure in STEM gateway courses. The Center for Assessment, Design, Research and Evaluation (CADRE).

CADRE

PLEASE DIRECT ANY QUESTIONS ABOUT THIS REPORT TO:

Jessica.Alzen@Colorado.Edu

The relationship between the Learning Assistant Model and failure in STEM gateway courses

Introduction

Science, Technology, Engineering, and Mathematics (STEM) departments at institutes of higher education frequently offer their introductory courses in large, lecture-style classes that can serve over one thousand students per semester. This method of course offering is efficient and cost-effective. Due to the large number of students in these courses, however, the instructional approach is often lecture-based with little interaction between the instructor and students or among students (Mason & Verdel, 2001; Talbot, Hartley, Marzetta & Wee, 2015). In addition, these courses tend to have relatively high failure rates compared to other common first-year courses (Benford & Gess-Newsome, 2006; Twigg, 2003). As a result of these conditions, many students who begin their undergraduate careers with an interest in pursuing a STEM major either switch majors or drop out of college or university without a degree (Crisp, Nora, & Taggert, 2009; Gainen, 1995).

The Learning Assistant (LA) model was established at the University of Colorado Boulder (CU Boulder) in 2001 in response to high failure rates in these courses and other educational issues such as facilitating instructional innovation. Funded by a grant from the National Science Foundation, one goal of the LA program is to assist faculty members in transforming their pedagogy to research-based, conceptual, and learner-centered instruction (Otero, 2015). An example of the instructional innovations facilitated by the LA program is the use of undergraduate Learning Assistants (LAs). Faculty members recruit students from math and science majors who have an interest in teaching or who faculty think would make good LAs. The student LAs get a monthly stipend for working ten hours per week. They also receive training in teaching and learning theories by enrolling in a math and science education seminar taught by discipline-based education researchers. The role of LAs within a gateway STEM course is to facilitate group discussions among students, while focusing on developing conceptual understanding of the content. LAs focus on eliciting student thinking and helping students articulate and defend their ideas to others. This is typically done both during the larger lecture section of the course as well as during smaller meetings after the weekly lectures, often referred to as “recitation.” In addition, LAs meet with faculty members once per week to develop deeper understanding of the content, share insights about student learning, and prepare for future class meetings (Otero, 2015).

Undergraduate LAs play an important role in this model of instructional innovation that is notably different from more traditional graduate student teaching assistants (TAs). While the primary purpose of TAs is to support instructors with teaching, LAs focus on assisting students with learning. Further, as Talbot et al. (2015) point out, “LAs do not grade or have input in evaluating students and are therefore meant to be in a more ‘trusted’ position” (p. 28). LAs serve as “near peers” to the undergraduates in these STEM gateway courses as they are much closer to the student experience than graduate-level TAs. Researchers hypothesize that connections with LAs help students gain content knowledge through the pedagogical approaches employed by the LAs and also see themselves as capable of mastering the content because they see peers not unlike themselves helping to provide content (Otero, 2015; Talbot et al. 2015).

Prior research suggests that failing STEM gateway courses is one way by which students fail to persist to graduation in college. The LA program began, in part, in an attempt to help assuage this issue. The current study has two aims: 1) to better understand the relationship between the LA program and course failure rates in a local context (CU Boulder), and 2) to serve as a model for how directors of LA programs at other institutions might study the same relationship at their own institutions.

Literature Review

A major purpose of the LA program is to help faculty innovate instruction in STEM courses. The goas is to transition courses from traditional lecture-style instruction, where students passively listen to the instructor, to learner-centered instruction in which the goal is to engage learners in the active construction of knowledge that leads to conceptual knowledge gain (Otero, 2015; Talbot et al., 2015). The LA model facilitates innovation in a way that allows for students to have more frequent interaction with a content expert, thus increasing not only their own understanding, but also their experience with someone they see as knowledgeable in the field in which they are majoring (Otero, 2015; Talbot et al., 2015).

Although we expect the instructional innovations facilitated by the LA program to improve course failure rates, there is no prior research regarding the connection between exposure to the LA program and course failure. Instead, the body of literature regarding the LA program focuses on specific student assessments. For example, Pollock (2006) provided evidence regarding the relationship between instructional innovation including LAs and discipline-based assessments in introductory physics courses at CU Boulder by comparing three different introductory physics course models.

In the first model, trained TAs and LAs facilitated small group work in recitation sections. Students worked materials found in the University of Washington Physics Tutorials curriculum (McDermott & Shaffer, 2002). TAs and LAs did not provide answers to the homework as much as guided discussion through questioning techniques to help students construct their own knowledge via discussion. In the second model, TAs facilitated small group work during recitation in which students completed exercises in a Physics for Scientists and Engineers workbook (Knight, 2004) for half of the term. During the last half of the semester recitation was used to review homework in a more traditional fashion, with TAs directly answering questions from the homework assignments. In the third and final model, there was no small group work during recitation. Instead this time was used by TAs to provide answers to homework questions.

Pollock reported two sources of evidence related to student outcomes and the relative effectiveness of these three course models. First, he discussed average normalized learning gains on the force and motion concept evaluation (FMCE; Thornton & Sokoloff, 1998) generally. The approach using tutorials with LAs saw a normalized learning gain[1] of 66% on the FMCE from pre-test to post-test. Average learning gains for the approach using Knight's (2004) workbooks with TAs were about 59% and average normalized learning gains for the traditional approach were about 45%. Second, Pollock further investigated the impact of the different course implementations on higher and lower achieving students on FMCE scores. To do this, he considered students with high pretest scores (those with pretest scores >50%) and students with low pretest scores (those with pretest scores <15%). For both groups of students, the course implementation that included

[1] Normalized learning gains are calculated by finding the difference in average post-test and pre-test in a class and dividing that value by the difference between 100 and the average pre-test score. It is conceptualized as the amount the students learned divided by the amount they could have learned (Hake 1998).

recitation facilitated by trained TAs and LAs had the highest normalized learning gains as measured by the FMCE. Prior research suggests that traditional instructional strategies yield an average normalized learning gain of about 15% and research-based instructional methods such as active learning and collaborative learning yield on average about 63% average normalized learning gains (Thornton, Kuhl, Cummings, & Marx, 2009). We see that the average normalized learning gains for all three methods to be much higher than what the literature would expect from traditional instruction but that the course model including LAs is aligned with what is expected from research-based instructional strategies. The FMCE is a concept inventory commonly used in undergraduate physics education to provide information about student learning on the topics of force and motion—a key concept in introductory physics. Since the FMCE does not include all of the key topics in introductory physics, it is not a direct indication of student grades in a course. However, it does provide information regarding student learning on some important concepts in these introductory courses (Thornton & Sokoloff, 1998).

Pollock's (2006) approach to understanding the impact of these instructional innovations and tutorials with the use of LAs provides some evidence about the relationship between courses including the LA experience and student outcomes. However, there are some key limitations to this work. This study used observational data. Students did not know prior to enrollment which instructional approach would be enacted in any course, so there is no bias due to students choosing a particular method of instruction. However, the study took place over multiple semesters with multiple instructors, and none of these differences were accounted for in the analysis. Although Pollock (2006) does not control for potential differences in the groups of students, he does provide evidence that pre-test scores for the three course models were relatively similar, suggesting that there were not large differences in the groups that might be the cause of the different gains seen on the FMCE. Despite this, there might still be other differences between the groups of students. The most obvious differences are that the instructor was different for each course model, that the curriculum changed across the different models, and that the courses likely included different cohorts of students who may have differed from one another in meaningful ways. Confounding variables such as these make it impossible to unambiguously attribute the differential gains to the LA or TA programs.

In a similar study at Florida International University (FIU) Goertzen, Brewe, Karmer, Wells, and Jones (2011) also investigated the influence of the instructional innovations made possible by the LA program in introductory physics. In contrast to Pollock's (2006) study that included University of

Washington Tutorials (McDermott & Shaffer, 2002) and Physics for Scientists and Engineers workbook (Knight, 2004), FIU used Open Source Tutorials (Elby, Scherr, Goertzen, & Conlin, 2008) developed at University of Maryland, College Park as the curriculum for the transformed lab courses. Goertzen et al, used the Force Concept Inventory (FCI; Hestenes, Wells, & Swackhamer, 1992) as the outcome of interest in their study. They found that those students exposed to labs transformed by the LA model had a 0.24 increase in mean raw gain in scores from pre-test to post-test while students in classes that were not transformed only saw raw gains of 0.16. They report this translates to an effect size of 0.59. Similar to the FCME, the FCI assesses student understanding of concepts related to force—another key concept in introductory physics. Just as with the FCME, the FCI is not a direct indication of student grades in a course as it only concerns a specific concept. However, taken together with Pollock's (2006) research, the increased assessment scores for both assessments and both institutions suggest that there may be a relationship between exposure to the LA program and student grades.

The previous studies each only considered one assessment of physics learning at individual institutions. In a more comprehensive study, White, Van Dusen, and Roulades (2016) conducted an investigation of the impacts of the LA model on student learning in physics across institutions with multiple assessments. In their study, White et al. used paired pre/post-tests from four concept inventories (FCI, FMCE, BEMA, and Conceptual Survey of Electricity and Magnetism [CSEM]) at 17 different institutions. Researchers used data contributed to the Learning Assistant Alliance through their online assessment tool, Learning About STEM Student Outcomes[2] (LASSO). This platform allows for institutions to administer several common concept inventories, with data securely stored on a central database to make investigation across institutions possible (Learning Assistant Alliance, 2018). In order to identify differences in learning gains for students who did and did not receive LA support, White et al. tested differences in course mean effect sizes between the two groups using a two-sample t-test. Across all of the concept inventories, White et al. found average Cohen's d effect sizes 1.4 times higher for LA-supported courses compared to courses that did not receive LA support. These findings across assessments suggests consistent increased learning on key concepts in introductory physics.

[2] More information about joining LASSO and resources available to support LA programs, visit https://www.learningassistantalliance.org/

There is also work that focuses on how the LA program affects student outcomes in other STEM fields. Talbot et al. (2015) focused on how transforming STEM gateway courses in biology might influence student satisfaction with an introductory Biology course. Although the focus of this study was primarily on student satisfaction survey data, Talbot et al. also provided minimal analysis with student outcome data. The researchers collected data from the Conceptual Inventory of Natural Selection (CINS; Anderson, Fisher, & Norman, 2002) in two sections of General Biology II, one with and one without LA support. A different instructor taught each section, but both the instructor and the course received relatively similar course ratings. Talbot et al. reported that the average normalized learning gain in the course with no LAs was -0.08 while the gain in the section with LAs was 0.49. Despite these encouraging results, the study design did not control for potential confounding factors that may have also influenced learning gains on the CINS. Similar to the other studies presented, it is possible that the differences in learning gains between LA and non-LA supported courses might have been due to other factors than the LA support itself such as difference in implementation of instruction or difference in class composition. However, this study does provide more consistent evidence regarding a relationship between exposure to the LA program and higher scores on assessments of key course concepts.

Although there has been some research on the relationship between the LA program and course-related outcomes, no prior research attempts to examine the relationship between taking LA-supported courses student grades directly. Further, none of this prior research considers student outcomes while controlling for variables that may confound this relationship. This study thus represents an extension of the previous work both in terms of the outcome of interest and the methodology.

Data

Data for this study comes from administrative records at CU Boulder. We focus on students who entered the university as full-time students for the first time in the fall semester from 2001 – 2016 overall. Student-level data for every department includes university enrollment term, term in which the course was taken, grade in the course, race/ethnicity, gender, admissions test scores (SAT or ACT), transfer student status, first generation status, an indicator for whether a student ever received financial aid while enrolled at CU Boulder, and high school GPA. In addition to administrative data, we also have course-level variables. The courses included in this analysis are as

CADRE

follows: Calculus I and II for Engineers in applied math (APPM), General Chemistry I and II in chemistry (CHEM), Calculus I and II in math (MATH), and Physics I and II in physics (PHYS). We focus on these particular courses because each is an introductory course in its respective department. As discussed in the literature review, a relatively high number of students often fail these classes, and failure in these courses is what often prohibits students from pursuing majors in these STEM fields. From 2001-2016, the average failure rate across these courses was about 15%. This is consistent with the failure rates in such courses at research institutions documented in previous literature, but failure rates for these courses is often closer to 30-40% at comprehensive universities and as high as 50-60% at community colleges (Twigg, 2003). Variables at the course level include instructor and whether the particular section of the course included LA support or not.

Instructors play a salient part in whether students fail a course or not. Thus, we feel including instructor effects to be one of the most important variables in this analysis. As a result, we include only instructors who have taught with and without LA support so that we only include instructors who are invested in instructional innovations made possible by the LA model. For APPM, CHEM, and PHYS, we limit the sample to only include course sections whose instructors who taught at least 100 students with and 100 students without LA support. Due to the nature of introductory courses in MATH described below, we limit the data in that department to include instructors who taught at least 40 students with and 40 students without LA support in the introductory courses.

The timing and nature of LA program adoption in each of the departments represented in our analysis is quite varied. Thus, the data used for each department is slightly different. Further, understanding of the method of LA program implementation in each department is important for proper consideration of the results from the following analysis. We next turn to a careful description of the department-specific data as well as the respective instantiations of the LA program at CU Boulder.

Applied Math (APPM)

Calculus I and II in APPM include three weekly lecture meetings, lasting 50 minutes each. Additionally, a 1-unit course called “workgroup” is an option for all Calculus I/II students in this department. Workgroup is an extra 90-minute session in which small groups of 4-6 students work on conceptual problems related to the weekly course material. LAs began facilitating these workgroups in 2004, and hence data from APPM includes first-time freshmen who enrolled at the university between 2004 and 2016. Signing up for a workgroup is voluntary, and hence students in every section of Calculus I and II in APPM have the opportunity to be exposed to LA support. However, those students who are expected to struggle more by their advisors or instructors are strongly encouraged to sign up for the workgroup. Additionally, instructors suggest that students who are not confident in their math abilities should sign up for the workgroup. In contrast to the other departments in this study, APPM does not use LAs during regular course meetings or during a required recitation meeting. Instead, LAs are only used in the optional workgroups. However, these LAs still participate in the weekly prep meetings with instructors and are still required to take the pedagogy course.

There are two selection bias issues in APPM. First is that of time on task. Students who sign up for the workgroup not only gain exposure to LA support, but they also gain an additional 90 minutes of time each week formally engaging in calculus material. There is no way to separate the effects of this additional time spent doing calculus each week from exposure to the support of undergraduate LAs. This selection bias might lead us to overestimate the relationship between LA exposure and course failure. Although we may not be able to disentangle whether it is the additional time on task or the presence of the LAs themselves, it is important to note that the LA model for instructional innovation in workgroup is what allows for the additional time on task to take place. Additionally, the method by which students are selected to receive LA support in this department is also problematic for disentangling the unique relationship between LAs and course exposure. Although all students have the opportunity to sign up for the additional workgroup, the fact remains that students who are expected to struggle or those who feel less confident are more strongly encouraged to sign up for the extra support. This difference in students who are encouraged to sign up for workgroup from those who are not might lead to underestimating the relationship between exposure to LA support and course failure.

Chemistry (CHEM)

Data for CHEM includes first time freshmen who enrolled at the university from 2006-2016. We limit to this subset of data because 2006 is the year in which CHEM started using LAs. In this department, LAs attend the weekly lecture meetings and assist small groups of students during activities such as answering "clicker" questions. Instructors present questions designed to elicit student levels of conceptual understanding. The questions are presented to the students, they discuss the questions in groups, and then respond using individual "clickers" based on their selection from one of several multiple-choice options. LAs help students to think about and answer these questions in the large lecture meetings. In addition, every student enrolled in General Chemistry I and II is also enrolled in a recitation section. Recitations are smaller group meetings of approximately 30 students. In these recitation sections, LAs facilitate small group activities related to the weekly lecture material for 3-5 students. The materials for these recitation sections are created by the lead instructor for the course and are designed to help students investigate common areas of confusion related to the weekly material.

A unique aspect of LA support in CHEM is the method by which students received LA support. From 2008 – 2013, LA support was only offered in the "on semester" sections of chemistry. "On semester" indicates General Chemistry I in the fall and General Chemistry II in the spring. Thus, there were few opportunities for those students who took the sequence in the "off semester", or General Chemistry I in the spring and General Chemistry II in the fall to receive LA support in these courses during the span of time covered in this analysis. The most typical reason why students take classes in the "off semester" are that they simply prioritize other courses more in the fall semester, so there is insufficient space to take General Chemistry I; they do not feel prepared for General Chemistry I in the fall and take a more introductory chemistry class first; or they fail General Chemistry I the first time in the fall and re-take General Chemistry I in the spring. This method of assignment to receiving LA support may overestimate the relationship between receiving LA support and course failure in this department.

Mathematics (MATH)

Data for MATH includes first time freshmen who enrolled at the university from 2005 – 2013. We make this data limitation because unlike the first two departments who have students receiving and not receiving LA support in each course every year, MATH switched from having no sections of Calculus I and II with LAs to offering the courses with LAs in every section in every semester. This shift occurred in Calculus I in Fall 2008 and in Calculus II in Spring 2010. We limit our data to 2005 – 2013 to include a roughly balanced timeframe from before and after the change to full LA support in this department. Something else different in MATH is the fact that although Calculus I and II are still seen as the gateway courses to pursuing a math major, this department eliminated large gateway courses. Instead, each section of Calculus I and II has fewer than 40 students each semester. Students participate in weekly lectures as well as an additional meeting facilitated by LAs and a TA in these courses. During this weekly meeting, students work in small groups to complete carefully constructed activities designed to enhance conceptual understanding of the materials covered during the weekly lecture. The LAs and the TA assist student learning as they complete these activities once a week, but are not present during the regular lecture meeting.

Physics (PHYS)

Data for PHYS includes first time freshmen who enrolled at the university from 2001 – 2010. Like MATH, PHYS either offered introductory physics with or without LAs in any given semester. Physics I was offered only with LA support during the 2003-2004 school year and then from 2006 to the present. Physics II shifted to include LAs in every section of the course in every term in Fall 2004. Similar to MATH, we include data from 2001 – 2010 in an effort to include relatively balanced data from the time before and the time after the switch to full LA support occurred.

An additional anomaly in PHYS is that the switch to the LA model happened concurrently with the adoption of the University of Washington Tutorials in introductory physics (McDermott & Shaffer, 2002). LAs facilitate small group work with the materials in the University of Washington Tutorials during recitation meetings. In other words, it is not possible to separate the effects of the content presentation in the Tutorials from the LAs facilitating the learning of the content. The results presented here may overestimate the relationship between receiving LA support and failure rates in introductory physics. However, it should be noted that the University of Washington Tutorials require a low student-teacher ratio. Proper implementation of this curricula is not possible without the undergraduate LAs helping to make that ratio possible.

In both PHYS and MATH there is also an historic threat to validity. That is, the results of the current analysis are threatened by any historic changes that occurred over time such as changes in curriculum or faculty, or even just historical differences in cohorts of students. As an example, if we observe that failure rates for students in MATH and PHYS introductory courses decreased after the switch to providing LA support, one interpretation would be that the LA program decreases failure rates. However, we could not rule out the possibility that failure rates decreased due to other factors that also changed over time. It could also be that the university implemented other student supports at the same time, that the types of students recruited to and accepted into MATH and PHYS changed, or that the external social environment encouraged more students to persist in school. There is no way to determine conclusively which of these (or other) factors may have caused the changes in failure rates. Conversely, it could also be true that if failure rates increased after switching to exclusively providing LA support, it might be that failure rates changed because of factors such as less-promising students were accepted into these departments or social climate on campus discouraged persistence in college. As a result, there is no way to clearly know in which direction and to what extent the relationship between LA support and course failure in these departments might be obscured due to historical confounds for which we are unable to control.

Outcome of interest

Although each department uses a slightly different dataset, the outcome of interest is consistent. The purpose of this study is to answer the following question: How do failure rates compare for students who do and do not receive LA support in each STEM gateway courses? The outcome of interest, receiving a final grade of a D or an F or W (based on withdrawal from the course), is a binary variable: a student either fails a course (receives a D/F/W), or the student does not. The comparison of interest in this study are the failure rates for those who took any of the gateway courses described above with LA support compared to those who took the same courses without LA support. Ideally, we would design a controlled experiment to estimate the causal effect of LA exposure on the probability of failing. To do this, we would need two groups of students: first, those who were exposed to LA support in a STEM gateway course, and second, a comparable group that significantly differed only in that they were not exposed to LA support in any STEM gateway course. However, many institutions, including CU Boulder, do not begin their LA programs with such studies in mind, the available data do not come from a controlled experiment. Instead, we must rely on observational data.

As described above, assignment to receiving LA support is not random in any of the four departments considered in this study, so students who received LA support in their gateway courses likely differed from students who do not receive LA support in many ways. All of the confounds described above make it difficult to determine whether we might over or under estimate the effect of the LA program on failure rates. For example, we might overestimate the effect if the university also pursued other student supports that decreased failure rates that similarly influenced those students who received LA support. This seems plausible given university-level goals for student success. Alternatively, we might also under estimate the effect if the students who struggle the most receive LA support. In other words, if the students who receive the LA support would have had higher failure rates to begin with, the magnitude of the effect might be masked. Further, the ways in which students find themselves receiving LA support in CHEM and MATH suggest that these students potentially differ on factors such as confidence in their abilities or in their abilities themselves. Prior research suggests that academic self-confidence and general self-efficacy have a stronger relationship to persistence in college than high school GPA and admissions test scores (Bean & Eaton, 2001; Lotkowski, Robbins, & Noeth, 2004). Unfortunately, our data set is limited to a small number of demographic and administrative variables including gender, race/ethnicity and whether a student is a first-generation college student or received financial aid. Additional academic achievement variables in our data include number of credits upon enrollment, high school GPA, and admissions test scores.[3,4] What this means is that our analysis controls for some differences between students that might also influence their probability of failing a course. However, we know from the onset of this analysis that we are unable to control for some important factors that we know are related to probability of failing a course as well as the mechanism by which students receive LA support at least in some departments. Although we do not have a controlled experiment that warrants causal claims, we desire to estimate a causal effect. The current study includes a control group, but it is not ideal because of the potential selection bias in each department described above. Our analytic approach is to control for some sources of selection bias, but we are limited by the availability of observed covariates. Thus, the results presented here lie somewhere between "true" causal effects and correlations.

[3] Since not all students take both the ACT and SAT, we use an SAT to ACT concordance table provided by the College Board to place all admissions tests scores on a comparable metric (College Board, 2016). We take combined SAT Critical Reading and Mathematics scores for students with SAT data and no ACT scores and use the concordance tables to translate SAT total scores to ACT Composite Scores. These scores range from one to thirty-six with a mean of about twenty-seven and a standard deviation of about four. We exclude students with no admissions test scores (about 5% of the population) from the analysis.

[4] All continuous control variables were standardized to have a mean of 0 and a standard deviation of 1 for ease of interpretation

Methods

We begin our analysis with a report of the raw failure rates for those who did and did not receive LA support in their STEM gateway courses. This is to provide a baseline description of failure rates in the courses of interest. We then move on to describe the differences between the two groups of students with respect to the demographic and administrative variables made available to us. Finally, we use the limited data we have in a logistic regression model in an attempt to disentangle some of the relationship between receiving LA support in gateway courses and probability of failing those same courses. The logistic regression models control for student demographic variables (entry term, gender, race/ethnicity, first-generation status, if a student received financial aid, number of credits at entry, high school GPA, and admissions test score). We also include dummy variables for course instructor and entry term. We include demographic variables to try to make a fairer comparison between LA and non-LA supported courses and to adjust for the potential confounds mentioned above.

Results

Raw failure rates for those students who do and do not receive LA support in their STEM gateway courses at CU Boulder are presented in Table 1. This table includes the number of students who enrolled in each of the courses included in this analysis as well as the number who failed based on exposure or lack thereof to LA support in the course.

The raw failure rates for these courses do not provide a clear picture about potential differences between failure rates when students are and are not provided LA support in their STEM gateway course. Students who received LA support in Calculus I and II in APPM had consistently lower failure rates. In CHEM, a higher percentage of students who received LA support failed General Chemistry I, but a lower percentage of students failed in Chemistry II with LA support. Conversely, in MATH, students who received LA support were less likely to fail Calculus I but more likely to fail Calculus II. Finally, in PHYS, students who received LA support were slightly more likely to fail Physics I and slightly less likely to fail Physics II.

Table 1. Raw Course Failure Rates by Department

APPM						
	Calculus I for Engineers			Calculus II for Engineers		
	# enrolled	# fail	% fail	# enrolled	# fail	% fail
LA	716	108	15	741	69	9
No-LA	2789	511	18	2828	445	16
Difference			-3			-7
CHEM						
	General Chemistry I			General Chemistry II		
	# enrolled	# fail	% fail	# enrolled	# fail	% fail
LA	4229	720	17	2620	247	9
No-LA	942	105	11	1029	141	14
Difference			6			-5
MATH						
	Calculus I			Calculus II		
	# enrolled	# fail	% fail	# enrolled	# fail	% fail
LA	328	60	18	96	22	23
No-LA	233	62	27	62	12	19
Difference			-9			4
PHYS						
	Physics I			Physics II		
	# enrolled	# fail	% fail	# enrolled	# fail	% fail
LA	2611	278	11	955	78	8
No-LA	766	79	10	315	31	10
Difference			1			-2

We next turn to student demographics in each department in Table 2. The purpose of investigating descriptive statistics for student characteristics is to try and understand potential differences between the students who did and did not receive LA support in STEM gateway courses that might also be related to differences in course failure rates.

CADRE

Table 2. Descriptive Statistics for LA and non-LA students

	LA	Non-LA	p-value
Applied Math	%	%	
Female	37	20	<0.01
Nonwhite	32	27	<0.01
First Generation	18	15	0.03
Receiving financial aid	47	44	0.06
	Mean (SD)	Mean (SD)	
Num. Credits at entry	10 (12)	10 (12)	0.27
High School GPA	3.82 (0.25)	3.75 (0.30)	<0.01
Test Score	28 (3)	28 (3)	0.04
N	1274	4696	
Chemistry	%	%	
Female	50	54	<0.01
Nonwhite	28	26	0.31
First Generation	17	18	0.44
Receiving financial aid	48	50	0.15
	Mean (SD)	Mean (SD)	
Num. Credits at entry	8 (11)	8 (11)	0.85
High School GPA	3.65 (0.34)	3.66 (0.34)	0.67
Test Score	27 (4)	26 (4)	0.06
N	5522	1794	
Math	%	%	
Female	45	47	0.56
Nonwhite	25	23	0.53
First Generation	15	21	0.06
Receiving financial aid	48	54	0.13
	Mean (SD)	Mean (SD)	
Num. Credits at entry	6 (9)	4 (7)	0.06
High School GPA	3.60 (0.36)	3.61 (0.36)	0.61
Test Score	27 (3)	26 (4)	0.05
N	421	295	
Physics	%	%	
Female	25	23	0.23
Nonwhite	23	23	0.70
First Generation	15	15	0.73
Receiving financial aid	51	52	0.80
	Mean (SD)	Mean (SD)	
Num. Credits at entry	9 (12)	7 (10)	<0.01
High School GPA	3.67 (0.33)	3.63 (0.34)	<0.01
Test Score	28 (3)	27 (3)	0.02
N	3006	924	

Similar to the raw failure rate data, the differences in student demographics vary by department. There is a statistically significant difference between nearly all the demographic variables for students in APPM who did and did not receive LA support in their gateway courses. This is potentially related to the fact that a certain kind of student is encouraged to sign up for LA support in this department. For example, the students who received LA support were more likely to be female, nonwhite, first-generation students who received financial aid. Students such as these are traditionally underrepresented in the sciences, so it is possible that these students tended to have lower self-confidence in their Calculus abilities. Conversely, students in CHEM only varied based on proportion of female students, and students in MATH only varied significantly based on admissions test score. Finally, students in PHYS varied based on number of credits at entry, high school GPA, and admissions test scores. Fewer differences between the groups of students in CHEM, MATH, and PHYS may be due to the fact that the ways students received LA support in these departments was not directly related to student feelings of self-efficacy. Among just these few variables, there is at least one statistically significant difference between those students who received LA support in their gateway courses and those who did not. This suggests that at least some of the differences in graduation rates provided in Table 1 could be explained by factors other than exposure to LA support in STEM gateway courses. However, we anticipate the biggest differences to be apparent in APPM based on the number and magnitude of differences among demographic variables presented in Table 2 as well as the mechanism by which students in this department receive LA support.

Logistic regression

We next use logistic regression to control for the influence of potentially confounding variables. The results shown in Table 3 reflect two different models. The first only included control variables for instructor, course, LA exposure, and an interaction between LA exposure and the second course in the sequence. This last covariate allowed us to see if there were differential effects for the LA program in the first versus the second course in any given gateway course sequence. Model 2 included all of the same covariates as Model 1 plus the student demographic variables. Comparing the results from these two models allows us to understand the extent to which student demographic variables accounted for variation in the relationship between LA support and course failure. Although instructor (Models 1 & 2) and entry term (Model 2) fixed effects are controlled, the relevant estimates are excluded from Table 3 to simplify presentation of the estimates. The results are reported in odds ratios (with confidence intervals) for course failure for each control variable. An odds ratio represents the proportional change in the odds of an event occurring (here failing a course) for a one-unit increase in an independent variable. An odds ratio equal to 1.0 indicates that changes in the independent variable are not associated with changes in the odds of failing a course. Odds ratios less than 1.0 indicate that changes in the independent variable are associated with a decrease in the chances of failing, while odds ratios greater than 1.0 indicate an increased chance of failing. Confidence intervals that include 1.0 indicate an odds ratio that is not statistically significant.

Consider first the contrast between Model 1 and Model 2 across departments. Contrary to expected, controlling for student demographic variables did not make a huge impact on the coefficient estimate for LA support in APPM. This suggests that although we expect students in this department who receive LA support to have lower prior achievement and lower feelings of self-efficacy, controlling for the prior achievement variables we have available to us did not make a significant difference in the estimated effect of the LA program on failure rates in this department. The difference between Model 1 and Model 2 in the other departments is larger, but still relatively non-significant except for in the PHYS department. Here, we see an increase in the estimate relationship between LA support and course failure by a magnitude of about 0.7. However, the estimate is not statistically significant in either Model 1 or Model 2. Overall, what we see in comparing these two models is that it does not appear that the demographic variables available to us do much to account for differences among students related to the relationship between LA support and course failure.

Next, looking specifically at the estimates for the relationship between LA support and course failure in the first course in the series, we see inconsistent results across department. Table 3 shows that students participating in an LA-supported course are less likely to fail the course relative to students who took the course with the same instructor without LA support, conditional on student covariates in APPM, but more likely to fail in CHEM. Specifically, holding constant all other variables in APPM, students who had LA support in Calculus I have odds of failing that are 23% lower than those of students who did not receive LA support in the same course. Chemistry I students who receive LA support have 50% greater odds of failing. However, there is no statistically significant relationship between taking an LA-supported course and course failure in MATH or PHYA in Model 2.

We further investigated the relationship between LA exposure and course failure rate in the second course in each sequence (i.e. Calculus II, Chemistry II, and Physics II). As seen in Table 3, the difference in the relationship between LA support and course failure in the second course was only different from the first in CHEM. In order to find the magnitude of the relationship between LA support and General Chemistry II, we multiply the estimates for LA support in the first course, the second course, and LA*Second course (i.e. the interaction between the two. This product (1.50*1.26*0.49) is equal to 0.93. In other words, although we see a 50% increase in the odds of failing General Chemistry I when there is LA support, we see 7% lower odds of failing General Chemistry II with LA support.

Table 3. Logistic Regression Odds Ratios (Confidence Intervals) for Course Failure

	Applied Math		Chemistry		Math		Physics	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
LA support	**0.78**	**0.77**	**1.65**	**1.50**	**0.65**	0.72	1.08	1.83
in 1st course	**(0.62, 0.98)**	**(0.60, 0.99)**	**(1.33, 2.07)**	**(1.17, 1.95)**	**(0.43, 0.98)**	(0.38, 1.33)	(0.61, 1.60)	(0.99, 3.34)
Second	0.92	1.06	1.35	1.26	0.60	0.54	0.99	1.02
course	(0.80, 1.06)	(0.90, 1.25)	(0.99, 1.86)	(0.85, 1.87)	(0.28, 1.21)	(0.22, 1.21)	(0.79, 1.49)	(0.6, 1.69)
LA support*	0.70	0.80	0.38	0.49	1.73	2.52	0.76	0.83
2nd course	(0.49, 1.00)	(0.55, 1.17)	(0.27, 0.53)	(0.32, 0.75)	(0.69, 4.44)	(0.82, 7.97)	(0.44, 1.30)	(0.44, 1.59)
Female	---	1.02	---	0.99	---	1.07	---	**1.93**
		(0.85, 1.22)		(0.86, 1.14)		(0.7, 1.64)		**(1.51, 2.47)**
Nonwhite	---	0.99	---	1.08	---	1.09	---	1.00
		(0.84, 1.17)		(0.93, 1.26)		(0.67, 1.74)		(0.77, 1.28)
First	---	**1.28**	---	**1.36**	---	1.43	---	**1.46**
Generation		**(1.05, 1.55)**		**(1.14, 1.63)**		(0.84, 2.39)		**(1.1, 1.92)**
Receive	---	1.03	---	1.01	---	0.98	---	0.86
Financial Aid		(0.89, 1.20)		(0.88, 1.17)		(0.65, 1.49)		(0.69, 1.06)
Standardized	---	**0.75**	---	**0.87**	---	**0.83**	---	**0.78**
entry credits		**(0.68, 0.83)**		**(0.78, 0.95)**		**(0.63, 1.07)**		**(0.67, 0.91)**
Standardized	---	**0.63**	---	**0.55**	---	**0.55**	---	**0.57**
HS GPA		**(0.59, 0.68)**		**(0.51, 0.59)**		**(0.44, 0.68)**		**(0.51, 0.63)**
Standardized	---	**0.81**	---	**0.81**	---	0.83	---	**0.81**
Test Score		**(0.75, 0.88)**		**(0.75, 0.88)**		(0.67, 1.04)		**(0.72, 0.91)**
N	5700		6830		710		3922	

NOTE: Bold font indicates statistically significant estimates

Odds ratios are difficult to compare to the associations implied in Table 1 as those failure rates are expressed in terms of percentage of students who failed each course. To facilitate a comparison to Table 1, we express the results from the logistic regression in terms of the marginal difference in the probability of course failure for specific groups of students who did and did not receive LA support in their gateway courses. One group of students we might consider are those who entered college in 2010, took their gateway course from an instructor with an average failure rate are male, white, non-first-generation college students, who did not receive financial aid, had the mean number of credits at entry, mean HS GPA, mean admissions test scores, and are not exposed to LA-supported courses. These students take on values of 0 for all variables in the model except for the particular instructor and entry term. The change in probability of failing for these students who only differ in that they received LA support in their gateway courses compared to those who do not for both courses is presented in Table 4.

Table 4. Difference in probability of course failure

Department	Course 1	Course 2
APPM	-3	-5
Chemistry	3	-2
Math	-6	11
Physics	2	1

Note: Reference categories for contrasts above are male, white, non-first-generation college students, do not receive financial aid, have the mean number of credits at entry, mean HS GPA, mean admissions test scores, and are not exposed to LA-supported courses. These students entered CU Boulder in 2010 and took their introductory courses from an instructor with an average course failure rate.

The results in Table 4 indicate that the change in probability of failing a course[5] from the logistic regression is rather similar to the differences presented in Table 1. We see consistently lower failure rates in APPM, mixed results in CHEM and MATH depending on course, and consistently higher, albeit of modest magnitude, in PHYS. Something to note is that this adjusted association varies as a function of a student's high school GPA. For the same group of students with high school GPA one standard deviation below the mean, the difference in probability of failing when exposed to LA support is 1-3% lower, meaning there is a lower likelihood of failing for these students. In other words, students with lower high school GPAs who are exposed to the LA program are less likely to fail than those who are not. Conversely, the difference is 1-3% higher for those students with high school GPAs one standard deviation above the mean, or a slightly higher chance of failing. That is, those students with higher high school GPAs who are exposed to the LA program are slightly more likely to fail than those who are not. We interpret this to mean that the LA program

[5] The Appendix provides details regarding the process for converting the results from a logistic regression into probability estimates.

has a greater positive relationship with passing a class for students who have lower prior achievement.

Recall the earlier descriptions of the differential implementations of the LA program in each department. We hypothesized that we would overestimate the relationship between the LA program and course failure rates in APPM and CHEM due to the ways students were assigned to receive LA support, and we were unsure about the direction of potential misattribution in MATH and PHYS due to historical confounds. Our results are consistent with our hypotheses in all departments except for CHEM. Based on our hypotheses about CHEM, we would have expected to see results more similar to that in APPM. Although we see this in General Chemistry II, we do not see the same in General Chemistry I. More work is needed in CHEM to understand the students who do and do not receive LA support to further understand these results.

First-year analysis

Next, we consider sensitivity analyses to test the robustness of the findings from Model 2 in Table 3 by limiting the sample to those students who took the courses in Table 1 with or without LA support in their first year at CU Boulder in APPM and CHEM. Students more often take large gateway courses in their first year on campus as opposed to later in their academic careers, so in many ways, these students are the target audience for the LA program. We limit the sample to only these two departments because most cohorts in MATH and PHYS only had the option of LA or not in their first year, so there is no variability in exposure to LA support available for these students. This analysis serves as a way of testing whether the relationship between having LA support in gateway courses and failing those courses is potentially different for those students who take the courses in their first year as students in APPM and CHEM. Odds ratios for exposure to LA support for this subset of students appear alongside the odds ratios for all students in Table 5.

Table 5. Logistic Regression Odds Ratios (Confidence Intervals)

	APPM		CHEM	
	All	Y1	All	Y1
LA support	**0.77**	0.82	**1.50**	**2.10**
	(0.60, 0.99)	(0.62, 1.08)	**(1.17, 1.95)**	**(1.34, 3.31)**
2^{nd} course	1.06	0.91	1.26	2.79
	(0.90, 1.25)	(0.75, 1.10)	(0.85, 1.87)	(0.74, 10.69)
LA support	0.80	0.80	0.49	0.32
*2^{nd} course	(0.55, 1.17)	(0.51, 1.24)	(0.32, 0.75)	(0.15, 0.71)
N	5700	5105	6830	3466

NOTE: Bold font indicates statistically significant estimates

Generally, we see similar trends for the relationship between LA support in gateway courses and course failure rates. Although the estimate for LA support in APPM is no longer statistically different from 1 in the first-year analysis, the magnitude of the estimate is relatively similar to that from the full sample. The CHEM estimates remain statistically significant and in the same direction, but the magnitudes of the estimates change in the first-year analysis. In general, this sensitivity analysis supports the initial estimates in these departments.

Course-level analysis

In one final investigation, we consider the relationship between LA support in a given section of a given course and the proportion of students who fail that course. In other words, we consider the same vein of analysis as conducted up until this point, but we consider it at the course-level rather than at the student level. Here, the outcome of interest is the proportion of students who fail a given course in a given term from a given professor. As a result, this is a linear regression as opposed to logistic. The results from a model weighted by sample size and including aggregated controls as well as fixed effects for instructor (not shown) appear in Table 6.

Table 6. Course-Level Linear Regression Estimates (Confidence Intervals) for Course Failure

	Applied Math	Chemistry	Math	Physics
LA support	<-0.01	0.01	-0.04	0.01
	(-0.01, 0.01)	(-0.03, 0.05)	(-0.13, 0.04)	(-0.03, 0.05)
Proportion Female	-0.03	0.07	0.30	0. 17
	(-0.20, 0.15)	(-0.41, 0.56)	(-0.10, 0.70)	(-0.22, 0.56)
Proportion Nonwhite	0.11	**0.54**	-0.33	0.27
	(-0.06, 0.27)	**(0.03, 1.04)**	(-0.71, 0.05)	(-0.14, 0.68)
Proportion First Generation	-0.13	-0.47	0.02	-0.21
	(-0.38, 0.12)	(-1.28, 0.34)	(-0.50, 0.55)	(-0.81, 0.40)
Proportion Receive Financial Aid	<0.01	0.04	0.22	0.03
	(-0.13, 0.14)	(-0.27, 0.35)	(-0.13, 0.57)	(-0.26, 0.32)
Mean standardized entry credits	0.01	-0.01	0.02	<0.01
	(-0.01, 0.02)	(-0.04, 0.02)	(-0.03, 0.07)	(-0.02, 0.03)
Mean standardized HS GPA	**-0.03**	-0.03	**-0.07**	**-0.03**
	(-0.05, -0.02)	(-0.07, 0.01)	**(-0.12, -0.02)**	**(-0.05, -0.01)**
Mean standardized Test Score	-0.02	0.02	-0.01	<-0.01
	(-0.04, 0.01)	(-0.01, 0.05)	(-0.06, 0.05)	(-0.02, 0.03)
N	128	42	39	41

NOTE: Bold font indicates statistically significant estimates

We see that having LA support has no statistically significant relationship on the proportion of students who fail the class holding all other variables constant in any department. This is similar to what was seen in MATH and PHYS in the previous analysis. However, the results are different in APPM and CHEM. Specifically, the student-level analysis suggested that LA support in APPM and CHEM courses were related to lower and higher probability of failing respectively, but the course-level analysis suggests no statistically significant relationship between LA exposure and proportion of students who fail. This difference is likely due to changes in sample size when considering the data at the course level. Nonetheless, more investigation into the LA program in these departments is still needed to better understand the relationship between exposure to the program and student-level outcomes in APPM and CHEM.

Conclusion and Limitations

In this study, we investigate the relationship between receiving LA support and course failure in STEM gateway courses at CU Boulder. The results here do not provide a coherent story about the relationship between the program and course failure rates across the four departments. However, this is not surprising given the individual adaptations of the LA program. Although we controlled for several student-level variables, it is still likely that we missed key variables that contribute to a student's propensity to fail a course. For example, a student's emotional health and student feelings of self-efficacy influence success in college and are likely related to if they received LA support at least in two departments in this study (Lotkowski, Robbins, & Noeth, 2004), and we are unable to control for any such factors.

The knowledge that the use of LAs and assignment to receive exposure to LAs is different across departments indicates a need for further research. It is necessary to understand whether it is reasonable to compare the relationship between receiving LA support and failure rates in gateway courses across these departments. Implementation of the LA program may be so different across departments that comparing the relationships to course failure rates across departments is like comparing apples to oranges. A more detailed understanding of the conditions under which students find themselves in the LA-supported versus non-LA-supported experiences in each department as well a deeper understanding of what the LA experience is like in each department are necessary in order to know how to best interpret the results of this study. In addition, future investigation should occur in programs at institutions that have stronger warrants for causal claims. For example, programs or institutions willing to use LAs in random recitation sections within course sections taught by the same instructors during the same semesters would mitigate confounding variables present in this study such as instructor, historical changes over time, and the issue of "on" versus "off" semester courses.

This study focuses on the presence of undergraduate LAs in key STEM gateway courses at CU Boulder. However, we remind the reader that the LA program is more than the undergraduate LAs and their direct effects on students. It is a model for institutional change, insofar as the presence of LAs in multiple courses and departments can lead to changes in values and practices at an institution. Thus, it may not make sense to assume that the physical presence of LAs is the only or even the most appropriate way to measure the impact of the LA program overall. It also may not make sense to attempt to isolate the presence of LAs from other course improvements. For example, although elements such as the additional time on task in APPM and the adoption of the University of Washington Tutorials materials in PHYS are listed as potential confounding variables in the current analysis, these changes are not independent of the LA program, and in some cases may actually have been implemented as a result of the LA program on the CU Boulder campus. In other words, any positive effects of increased time on task and the University of Washington materials could be attributed to the LA program, even if indirectly. Further, interpreting the relationship between LA support and course failures as causal assumes that having LA support for a single semester can influence student outcomes in that same semester, but it might also be the case that there are more long-term effects (c.f. Close, Mailloux-Huberdeau, Close, & Donnely, In press). Other research indicates a relationship between exposure to the LA program generally and lower failure rates in subsequent STEM gateway courses (Alzen, Langdon, & Otero, In press). These are all complications to understanding the effectiveness of the LA program and the potential impacts the program has on students' college experiences. Understanding the LA program in its entirety requires a much more comprehensive research approach than is presented here. There is a need for more focused, mixed methods work to better understand the extent of the influence the LA program has on a variety of student outcomes.

References

Alzen, J. A., Langdon, L., & Otero, V. K. (In press). The learning assistant model and DFW rates in introductory physics courses. Accepted to L. Ding, A. Traxler, & Y. Cao (Ed.s), A*IP Conference Proceedings: 2017 Physics Education Research Conference.*

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of research in science teaching*, *39*(10), 952-978.

Bean, J., & Eaton, S. B. (2001). The psychology underlying successful retention practices. *Journal of College Student Retention: Research, Theory & Practice*, *3*(1), 73-89.

Benford, R., & Gess-Newsome, J. (2006). Factors Affecting Student Academic Success in Gateway Courses at Northern Arizona University. *Online Submission*.

Close, E. W., Mailloux-Huberdeau, J.-M., Close, H. G., & Donnelly, D. (In press). Characterization of time scale for detecting impacts of reforms in an undergraduate physics program. Accepted to L. Ding, A. Traxler, & Y. Cao (Ed.s), A*IP Conference Proceedings: 2017 Physics Education Research Conference.*

College Board. (2016). Concordance Tables. Retrieved from https://collegereadiness.collegeboard.org/pdf/higher-ed-brief-sat-concordance.pdf

Crisp, G., Nora, A., & Taggart, A. (2009). Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a STEM degree: An analysis of students attending a Hispanic serving institution. *American Educational Research Journal*, *46*(4), 924-942.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical review special Topics-Physics education research*, *2*(1), 010105.

Elby, A., Scherr, R. E., Goertzen, R. M., & Conlin, L. (2008). Open-Source Tutorials in Physics Sense Making.

Gainen, J. (1995). Barriers to success in quantitative gatekeeper courses. *New Directions for Teaching and Learning*, *1995*(61), 5-14.

Goertzen, R. M., Brewe, E., Kramer, L. H., Wells, L., & Jones, D. (2011). Moving toward change: Institutionalizing reform through implementation of the Learning Assistant model and Open Source Tutorials. *Physical Review Special Topics-Physics Education Research*, *7*(2), 020105.

CADRE

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... & Wood, W. B. (2004). Scientific teaching. *Science*, *304*(5670), 521-522.

Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The physics teacher*, *30*(3), 159-166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141-158.

Knight, R. (2004). Physics for Scientists and Engineers. Essex: Pearson.

Learning Assistant Alliance. (2018). About LASSO. Retrieved from https://www.learningassistantalliance.org/public/lasso.php

Lotkowski, V. A., Robbins, S. B., & Noeth, R. J. (2004). The Role of Academic and Non-Academic Factors in Improving College Retention. ACT Policy Report. *American College Testing ACT Inc*.

Mason, D., & Verdel, E. (2001). Gateway to Success for At-Risk Students in a Large-Group Introductory Chemistry Class. *Journal of Chemistry Education*, *78*(2), 252–255. http://doi.org/10.1021/ed078p252

McDermott, L.C., and Shaffer, P.S. (2002). *Tutorials in introductory physics.*Upper Saddle Ridge, New Jersey: Prentice Hall.

Otero, V.K. (2015). Effective practices in preservice teacher education. In C. Sandifer and E. Brewe (Ed.), *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices* (pp.107-127). College Park, MD: American Physical Society.

Pollock, S. J. (2009). Longitudinal study of student conceptual understanding in electricity and magnetism. *Physical Review Special Topics-Physics Education Research*, *5*(2), 1-8.

Pollock, S. J. (2006). Transferring transformations: Learning gains, student attitudes, and the impacts of multiple instructors in large lecture courses. AIP Conference Proceedings, 818, 141–144. http://doi.org/10.1063/1.2177043.

Talbot, R. M., Hartley, L. M., Marzetta, K., and Wee, B. S. (2015). Transforming Undergraduate Science Education With Learning Assistants: Student Satisfaction in Large-Enrollment Courses. *Journal of College Science Teaching*, *44*(5), 24–30.

Thornton, R. K., Kuhl, D., Cummings, K., & Marx, J. (2009). Comparing the force and motion conceptual evaluation and the force concept inventory. *Physical Review Special Topics-Physics Education Research*, *5*(1), 010105.

CADRE

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, *66*(4), 338-352.

Twigg, C. A. (2003). Improving quality and reducing cost: Designs for effective learning. *Change: The Magazine of Higher Learning*, *35*(4), 22-29.

White, J. S. S., Van Dusen, B., & Roualdes, E. A. (2016). The Impacts of Learning Assistants on Student Learning of Physics. *arXiv preprint arXiv:1607.07469*.

CADRE

Appendix A: Converting log odds to odds ratios

Results from the logistic regression are reported as odds ratios in this report for ease of interpretation. However, most statistical software estimates and reports the results from logistic regression in log odds. This appendix provides the log odds estimates for APPM from Table 4 Model 2 in the current paper as well as an example of the calculations necessary to transform the output into probabilities as presented in Table 5.

Table A1. Logistic Regression Log Odds Estimates

	Applied Math
LA support	-0.26
Female	0.02
Nonwhite	-0.01
First Generation	0.25
Receiving Financial Aid	0.03
Standardized entry credits	-0.28
Standardized HS GPA	-0.45
Standardized Test Score	-0.21
Second course	0.06
LA support* Second course	-0.22
2010 Cohort	-2.06
Average Instructor	0.24
N	5700

Any log odds can be transformed into a probability through exponentiation. In this explanation, we consider the transformation for APPM students who entered in the 2010 cohort; had an average instructor; are white; male; non-first generation college students; who do not receive financial aid; and have an average value for number of credits at entry, high school GPA, and admissions test scores; and do not receive LA support. In order to find the marginal probability of failure for this type of student when exposed or not to LA support, we perform the following calculations.

$$P(\text{Grad6}=1) = \frac{e^{2010\ \text{Cohort}(1)+\text{Average Instructor}(1)+LA(1)}}{1+e^{2010\ \text{Cohort}(1)+\text{Average Instructor}(1)+LA(1)}} - \frac{e^{2010\ \text{Cohort}(1)+\text{Average Instructor}(1)+LA(0)}}{1+e^{2010\ \text{Cohort}(1)+\text{Average Instructor}(1)+LA(0)}} =$$

$$\frac{e^{-2.06(1)+0.24(1)-0.26(1)}}{1+e^{-2.06(1)+0.24(1)-0.26(1)}} - \frac{e^{-2.06(1)+0.24(1)-0.26(0)}}{1+e^{-2.06(1)+0.24(1)-0.26(0)}} =$$

$$= 0.11 – 0.14 = -0.03$$

This means that the difference in probability of failing Calculus I for the types of students described above who are instead of are not exposed to LA support is 3%. Although not shown, both of these calculations assume that all the other point estimates from the logistic regression are multiplied by 0 and thus not included in the calculation. The variables only appear in the calculation if there is an interest in understanding a contrast for a particular student profile that requires changing the value of the predictor from 0 to 1. Since fixed effects for cohort and instructor were used in the regression estimation, we must include a value for each variable. Hence the inclusion of the 2010 cohort and an anonymized instructor. If fixed effects had not been used, then an estimate for an intercept would have needed to be included in these calculations in place of the cohort and instructor values.

CADRE