

San Jose State University

SJSU ScholarWorks

Faculty Research, Scholarly, and Creative Activity

1-1-2025

Examining equity and graduation rates at two institutions using a course deficit model and the collaborative learning through active sense-making in physics curriculum

Cassandra A. Paul

San Jose State University, cassandra.paul@sjsu.edu

David J. Webb

University of California, Davis


Follow this and additional works at: https://scholarworks.sjsu.edu/faculty_rsca

Recommended Citation

Cassandra A. Paul and David J. Webb. "Examining equity and graduation rates at two institutions using a course deficit model and the collaborative learning through active sense-making in physics curriculum" *Physical Review Physics Education Research* (2025). <https://doi.org/10.1103/PhysRevPhysEducRes.21.010107>

This Article is brought to you for free and open access by SJSU ScholarWorks. It has been accepted for inclusion in Faculty Research, Scholarly, and Creative Activity by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Examining equity and graduation rates at two institutions using a course deficit model and the collaborative learning through active sense-making in physics curriculum

Cassandra A. Paul 

*Physics and Astronomy Department, Science Education Program, San Jose State University,
San Jose, California, USA*

David J. Webb 

*Department of Physics and Astronomy, University of California, Davis,
California, USA*



(Received 23 August 2024; accepted 10 December 2024; published 23 January 2025)

We have previously described the reformed introductory physics course, collaborative learning through active sense-making in physics (CLASP), for bioscience students at a large public research one university (Original University) and presented evidence that the course was more successful and more equitable than the course it replaced by several measures. Now we compare the original success of CLASP with an implementation at a second institution. We find that the original results hold at another institution despite some changes to the original curriculum and a somewhat different student population. We find that students who take CLASP are (1) less likely to drop, (2) less likely to fail, and (3) do as well in later coursework when compared to students who took the courses that CLASP replaced, even if that coursework is not similarly reformed. We find the above items to be independently true for historically marginalized students and remarkably, also find that (4) marginalized students who take CLASP are more likely to graduate from a science, technology, engineering, and mathematics field. We use a course deficit model perspective to examine these results and discuss some of the factors that may have contributed to this success. We argue that higher education has the tools needed to significantly increase equity and improve student success and retention.

DOI: [10.1103/PhysRevPhysEducRes.21.010107](https://doi.org/10.1103/PhysRevPhysEducRes.21.010107)

I. INTRODUCTION

A. Course reform

Retention in higher education has always been important, but it has taken on new significance now that enrollment in higher education is facing a pending enrollment cliff, which has been further impacted by the effects of COVID-19 pandemic and other factors involving the changing demographics of the USA [1–3]. One potential avenue for increasing retention is course reform, especially large introductory courses. In a large multi-institutional study, Hatfield *et al.* [4] find that when students struggle in one or more introductory STEM courses, it significantly reduces their chances of graduation. Because there is overwhelming evidence that students perform better when active learning techniques are used in STEM classrooms [5,6], reforming large undergraduate science, technology,

engineering, and mathematics (STEM) courses can provide a powerful pathway toward increasing retention in higher education.

Course reform has been one of the major areas of study in Physics Education Research (PER) over the long history of our field. Often the results from the reformed courses are compared to the those of the more traditional courses that the reformed courses replaced whereas side-by-side comparisons *across different institutions* are less common [7]. We know curricular success depends partly on context and that PER studies are primarily done at research 1 (commonly referred to as R1) institutions and so are not representative of the general college population [8]. This suggests that PER is in need of more replication studies at institutions that are significantly different from the institution of curricular origin.

Collaborative learning through active sense-making in physics (CLASP) is an active learning and model-based introductory physics reformed curriculum. When compared to the traditional course, CLASP has been shown to be correlated with increased student GPA and MCAT scores and resulted in high conceptual and attitudinal gains compared to national averages [9]. CLASP has also been shown to increase the odds of course success for a variety of

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

student demographics [10] so that overall success improved under CLASP in an equitable way. However, research on student success under this curriculum has been primarily limited to a single university.

In this paper, we present a side-by-side analysis of CLASP implementations at two very different institutions, 20 years apart. We show that there is a remarkable similarity across these schools in the ways that success in the course has become more equitable under CLASP.

B. Active learning

Over the past 30 years, there have been many physics course reform efforts (such as CLASP) featuring active learning via small-group peer-peer discussions as a major course component. In practice, these peer-peer discussions often take place in a studio-type setting where students work together in small groups on activities facilitated by the instructors [9,11–14]. The PER community and the greater Discipline-Based Education Research community have provided overwhelming evidence that active learning courses on average improve learning outcomes for students in these settings and beyond [5,6,15].

However, research has also shown that there is a lot of variety in student performance on various metrics between different demographic populations in active learning environments [11,12,16–19]. For example, in their 2014 paper, Eddy and Hogan discussed how different populations interact and perform differently in active learning classrooms suggesting that there is a more nuanced answer to the question, “Is active learning better?” [20]. They find that while all students performed better in an active learning environment, the gain in exam performance almost doubled for African-American students and improved similarly for the first generation college students. The African-American students had different perceptions of the class than other groups, suggesting that the mechanism for increased performance may not be the same for all groups of students. Therefore, it is important to consider for whom active learning is better, and by which metric. In this paper, we use several distinctly different metrics in order to develop a larger picture of equity issues and their changes under CLASP at these two institutions.

II. RESEARCH FRAMING

A. Course deficit model

In recent work [19], we have suggested that demographic differences in course grades are best understood using a course deficit model [21], rather than the more traditional student deficit model [22]. In this paper, we use an entirely different dataset and entirely different course changes in support of the same conclusion that a particular course structure may contain hidden bias against some demographic groups so that a course deficit model is the best way to understand differences in results between different

demographic groups. Using a course deficit model means that we attribute the differences in performance between different student groups in a given course to the features of the course itself [19,21]. The model is perhaps best understood in contrast to the student deficit model, which assumes that differences in course performance across different student demographic groups arise not from the course structure but from past inequities between different student demographic populations. Importantly, the course deficit model does not assume that students necessarily have equal backgrounds, it merely assumes that different student demographic groups are equally capable of succeeding in a given course and that any gaps in performance are instead the result of different classroom practices systematically favoring some populations over others.

B. Equity models

Since equity can mean different things to different people and in different contexts, we will be careful with our meaning in this paper. In this work, we apply two models of equity: equity of parity and equity of individuality [23]. Equity of parity works particularly well with a course deficit model, because in order for equity of parity to be achieved, all demographic groups must achieve the same outcome. Importantly, when equity of parity is used with a course deficit model, it does not ignore past inequities suffered by different demographic groups, it merely indicates that these inequities are inconsequential to success in the course. In this paper, we consider equity of parity achieved when the average success of individual demographic groups is statistically indistinguishable from the average of all other groups [23,24]. We also use the model of equity of individuality. This model of equity compares each group’s success to the prior success of that group instead of comparing it to a specific comparison group [23]. We will be careful to examine the impact of our interventions on many different student demographics in order to ensure that the intervention does not inadvertently disadvantage some groups even while increasing overall gains.

While it is customary in educational research to group all marginalized groups together into one single group for statistical purposes, this practice can hide important differences between demographic groups. Following the work of Shafer *et al.* [18] as well as our own [25], we recognize that different metrics yield different results among different student demographic groups. In our work in this paper, we are careful to examine the impacts of our intervention on many different student demographic groups. Because we want to ensure simplicity in our message, we have chosen to present marginalized student groups together when the outcome is similar for all groups. However, we include the comparisons by detailed student demographics in the Appendix for completeness.

III. RESEARCH QUESTIONS

In this paper, we examine several years of transcript data for students across different demographic groups in order to better understand how the switch from a traditional course to the active-learning-based collaborative learning through active sense-making in physics (CLASP) curriculum [9,10,26] has impacted different groups of students at two different institutions. We have previously shown that CLASP produces better learning gains on average than the traditional course it replaced at a large public R1 institution [9] and, in a conference proceeding, we also showed that the impact of the CLASP course at the original institution was positive across all ethnic demographic groups we examined [10]. Since CLASP was implemented several years ago at a second institution, this study is primarily a replication study of that conference proceeding. We show side-by-side results from both institutions.

As a part of a larger project in this study, we consider equity across student ethnicity by examining course grades and pathways toward graduation. We examine the success of students grouped by self-identified ethnicity (as defined and collected by each institution), comparing success in CLASP with that of the traditional course it replaced. The practice of examining improvements within specific student populations, without comparing their success to that of some normative group, is referred to as applying a model of equity of individuality [23]. We say more about how this model is used in our results and in our discussion.

We investigate the following research questions:

- (1) How do students of different demographic groups perform in the CLASP series as compared to the traditional course it replaced at each institution?
- (2) What evidence exists that CLASP students of different demographics are prepared for later coursework?
- (3) Are CLASP students of some demographics more likely to graduate than those who took the course that CLASP replaced?

IV. COURSE SETTING

As described in our previous work [9], the CLASP curriculum at the Original University (OU) was implemented in an introductory physics course required for many STEM degrees, including the biosciences. CLASP features an explicit focus on understanding models (including words, graphs, equations, etc.) and using these models in qualitative and quantitative analysis of real-world physical, chemical, and biological situations [9,26]. The active learning elements of the course are carried out in studio-type discussion/laboratory (DL) sections of 25–35 students. These DL sections include small-group activities where students work together to understand a model and practice applying it before engaging in whole-class discussions. At Original University, CLASP is a three-quarter (1 year)

series, meaning there are CLASP Physics A, B, and C courses.

At both Next and Original Universities, the students meet for 140 min twice a week to work through activity cycles of small group work and whole-class discussions. Lecture (during the time of the study on both campuses) was limited to one 80 min session per week. DLs are taught by either a faculty member or a graduate teaching assistant. At both locations, the lecture (typically taught by a faculty member) reviews or previews the topics discussed in the DL and does not introduce material not covered in the DLs. The DLs are considered the main driver of the curriculum. Exams and quizzes, which were conducted during the lecture section, determine most of the students' grades.

There are a few differences between the implementations of CLASP at the two universities. At Next University, the introductory series is also a year-long course but, since the Next University is on semesters, it is a two-semester series. An additional difference between the two universities is that at Next University only the A course is based on CLASP, the B course is a traditional course for the years in this study.

There are also differences in the order of the material in the two courses. Both universities (nontraditionally) begin their curricula with fundamentals concerning the conservation of energy including the three-phase model of matter, then introducing phase and thermal energies, then move into mechanical models of kinetic and potential energy. However, while the Original University covers particle models of bond and thermal energy and then moves into thermodynamics and the ideal gas model during the first quarter, Next University skips particle models and much of the thermodynamics, and instead moves to more traditional introductory physics concepts like conservation of momentum, Newton's laws, circular motion and kinematics. (Original University covers these topics in their B section of the course using the same materials.) These curricular ordering decisions were made so that the changes made to Next University's course did not change the content covered in either the A or the B section course.

Next University's discussion/laboratory (DL) sections are also generally smaller, with the maximum number of seats being 25 instead of Original University's maximum of 35. Finally, beginning in 2016, Next University also had undergraduate Learning Assistants in many DL and lecture sections.

V. METHODS

A. Grades and retention

Some studies have shown that grades are not necessarily good indicators of learning because students have an ability to solve physics problems without a good understanding of the physics content [27,28], therefore, student pass rates may not be a good indicator of learning goals. However,

grades still play an important role in the academic process. For example, because grades represent an official form of evaluation, good grades can encourage students, and poor grades can deter them [13]. Furthermore, pass rates are a very real indicator of college success, as course failures can mean students need to repeat coursework, drop out of their major, and/or leave college entirely.

While we will be using the absolute grade a student gets as one measure of success, we will not be using past grades as controls for any model that makes comparisons across demographics. It is common in quantitative PER for researchers to choose to control for students' prior preparation or grade point average. Because we are using a course deficit model, we do not intend to do this. This is for two reasons. First, as described earlier, the course deficit model ignores incoming differences and attributes any differences in course performance between student demographic groups to course structures. We expect to see differences in student preparation resulting from the structural racism evident in society. Thus, controlling for preparation could potentially hide the equity gaps we are trying to address [19,29]. Second, the grades that students receive in a given course have very real implications for students. Take this illustrative example: if one demographic group receives an average of D in a course, and another demographic group receives an average of B, but when we control for preparation then both groups perform equitably (as they both have the same expected growth), this hides the very different experiences and outcomes that each demographic group has in that classroom, and any resulting consequences of those differences (i.e., needing to retake the course, getting behind, and even dropping the major). Fortunately, the course deficit model provides the best explanation for differences in the success of different demographic groups so these problems with a student deficit model need not arise.

B. Original University context

Original University began offering trials of CLASP in 1995 and completely replaced the traditional lecture and lab course by academic year 1996–1997. This switchover creates the conditions of a natural experiment that allowed us to compare student successes in the traditional lecture/lab courses with successes in an active-learning studio-type courses (CLASP). For all analyses, we will use the same set of students at Original University that we used for our earlier paper [10], comparing those who took CLASP to those who took the traditional course CLASP replaced. The NonCLASP classes we include were offered from the Fall quarter of 1993 through the Summer of 1996 and the CLASP courses included were offered from Winter quarter of 1995 through Summer of 2000.

Ethnicity data are self-reported by all students from predetermined ethnicity categories created and named by Original University. Students were only allowed to select

TABLE I. Some demographics of our dataset from OU. The fractions of male students, HM students, and STEM majors in each set of either NonCLASP or CLASP courses is shown.

Group	N	Male	HM	STEM
NonCLASP	3386	0.44	0.12	0.90
CLASP	4849	0.38	0.13	0.93

one option but they could choose not to select any. The student population represented in the data is 84% US citizens and 16% permanent residents. Groups historically marginalized in physics [30] (we will use the acronym HM for this demographic grouping) in this study include African-Americans, Latina(o), Chicana(o) or Mexican American, Hispanic-other, Native American, and Pacific Islander students. The population fitting the HM classification is about 89% US citizens and 11% permanent residents.

The regular terms offered from Fall 1993 to Summer 2000 included 8235 students (of which we have ethnicity information for 7896). Table I gives a short demographic breakdown of these students over these years.

C. Next University context

Trial runs of the CLASP courses at the Next University (NU), following on the work at the original university (OU), were begun in 2013 and continued through 2014, running at the same time as the NonCLASP courses that would be replaced. Only the first semester (Physics A) of the two-semester series was changed to the CLASP format. To include enough students for analyses, we use these two years that had both versions being offered and also add in the attached 3 years, 2010–2012, of no CLASP courses and the 3 years, 2015–2017, of only CLASP courses.

Summer classes did not ever change to CLASP format but they make up only 3% of the total number of students which is not large enough to be a useful comparison group by themselves. Because these summer classes are also likely to include the highest percentages of nonstandard students (students from other universities, students trying to get out of academic probation, etc.), we will leave them out of all of our analyses and just compare the regular term CLASP with regular term NonCLASP students. This will change the numbers slightly but not any of our conclusions.

TABLE II. Some demographics of our dataset from NU. The fractions of male students, HM students, and STEM majors in each set of either NonCLASP or CLASP courses is shown.

Group	N	Male	HM	STEM
NonCLASP	1927	0.52	0.24	0.68
CLASP	1292	0.55	0.28	0.68

The regular terms offered from Spring 2010–Fall 2017 included 3219 students (of which we have ethnicity information for 2943). Table II gives a short demographic breakdown of these students over these years.

VI. DROPPING THE COURSE

A. Graphical examination of drop rates

To succeed in a class, the first thing a student enrolled in the class must do is to not withdraw from the class. So, first we will examine the fractions of students who dropped the two types of classes, how these fractions can depend on some demographic variables and compare the CLASP/NonCLASP results for the two very different universities separated by about 20 years in time. Figure 1 shows these drop fractions with a set of CLASP/NonCLASP comparisons for Next University on the left and for Original University on the right. To make the comparisons of equity across different demographic groups easier to do by eye, we have drawn a horizontal line through the overall drop fraction for CLASP course in red and NonCLASP courses in black. One sees that the CLASP courses at both universities (a) have considerably smaller fractions of students who choose to drop the course and that (b) these drop fractions have less of a dependence on demographic group than the NonCLASP courses. We examine these two findings statistically in the following section. It is also rather striking that these two results show up in a similar way for two very different universities 20 years apart. (For further graphical breakdown of demographic groups by gender and race/ethnicity, see Figs. 5 and 6 in Appendix D).

B. Statistical examination of drop rates

1. Drop-rate changes between NonCLASP and CLASP courses

For all students, the odds of dropping a NonCLASP course are significantly higher than the CLASP course at both universities. Table III shows these results along with the effect size of about 19 fewer CLASP students dropping per year and about 30 fewer CLASP students dropping per year at Original University.

In addition to finding fewer students dropping over all, we can use an equity of individuality model of equity, to examine the differences between CLASP and NonCLASP courses for *each* demographic group. To quantify the CLASP/NonCLASP differences we see in Fig. 1, we use logistic regression to estimate the ratio of the odds of a group member dropping a CLASP course to the odds of a member of the same group dropping a NonCLASP course. Equation (1) defines the odds of an event and our model is shown in Eq. (2). We use this model to fit our data separately for each demographic group:

$$\text{odds of dropping} = \frac{\text{drop probability}}{1 - \text{drop probability}}, \quad (1)$$

$$\text{Log}(\text{DropOdds}) = b_0 + b_{\text{CLASP}}\text{CLASP}, \quad (2)$$

where CLASP = 1 if the student was enrolled in a CLASP class and 0 if they were in a NonCLASP class. Equation (3) is used to convert a coefficient from the fit into an odds ratio

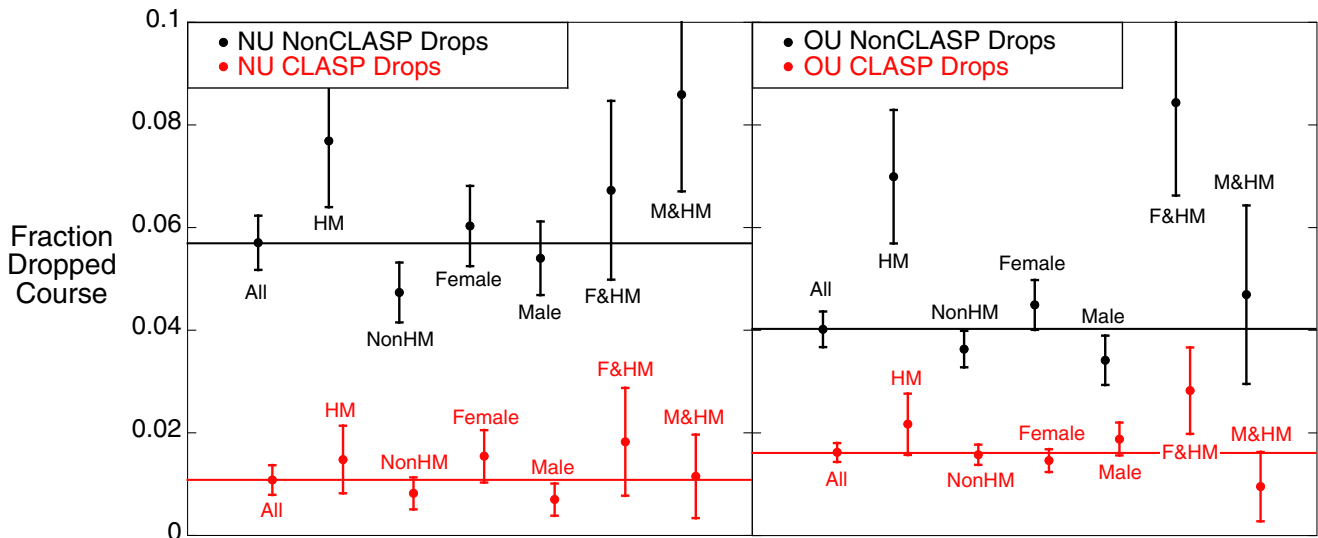


FIG. 1. The fraction of drops in NU CLASP and NonCLASP classes is shown on the left. Similar data for OU are shown on the right. The error bars are standard errors. In each case, several different ways of separating the students into demographic groups are shown. Specifically, two intersections of the other groups are also shown. For example, F and HM refer to the group of students who identified as female and also as members of racial or ethnic groups historically marginalized in physics.

TABLE III. Class-type comparisons of course dropping, for specific groups. For each university and for each demographic group in Fig. 1, we list the number of students in the measurement, the odds ratio for dropping = (odds of dropping a CLASP course)/(odds of dropping a NonCLASP course), the standard error for those odds ratios, the p -value for those odds ratios, and the estimated net number of students *per year* from that demographic group who would have dropped from the NonCLASP course but would benefit from the CLASP course by not dropping their class. The standard error on this number of students is in parentheses.

Group	NU				OU			
	N	Odds ratio (SE)	p -value	Extra NonCLASP yearly drops (SE)	N	Odds ratio (SE)	p -value	Extra NonCLASP yearly drops (SE)
All	3219	0.181 (0.052)	$<10^{-3}$	18.6 (2.4)	8235	0.404 (0.056)	$<10^{-3}$	29.5 (4.7)
HM	766	0.181 (0.088)	$<10^{-3}$	5.9 (1.4)	990	0.30 (0.10)	$<10^{-3}$	6.8 (2.0)
NonHM	2177	0.167 (0.067)	$<10^{-3}$	10.7 (1.8)	6906	0.425 (0.069)	$<10^{-3}$	21.0 (4.0)
Female	1510	0.245 (0.089)	$<10^{-3}$	8.5 (1.8)	4867	0.315 (0.059)	$<10^{-3}$	21.6 (3.7)
Male	1709	0.124 (0.058)	$<10^{-3}$	10.0 (1.7)	3304	0.569 (0.125)	$<10^{-3}$	6.9 (2.7)
F and HM	372	0.26 (0.17)	0.036	2.3 (0.9)	628	0.32 (0.12)	0.003	5.0 (1.8)
M and HM	394	0.124 (0.093)	0.005	3.7 (1.0)	362	0.20 (0.16)	$<10^{-3}$	1.9 (1.8)

$$\text{Odds ratio} = e^b = \frac{\text{CLASP DropOdds}}{\text{NonCLASP DropOdds}}. \quad (3)$$

Table III shows the odds ratios and standard errors for each fit of the model to measure the difference between CLASP classes and NonCLASP classes for each demographic group shown in Fig. 1. In addition, we have given the estimated net number of students per year (along with the standard error) who would have dropped if they were enrolled in a NonCLASP course but not if they were enrolled in a CLASP course. If CLASP and NonCLASP course produce equal odds of dropping the course, we would find an odds ratio = 1. As expected, all of these odds ratios are less than one with the smallest odds ratios across both universities for male students from historically marginalized groups showing that at NU they had eight times the odds of dropping NonCLASP as of dropping CLASP and at OU five times the odds of dropping NonCLASP compared to dropping CLASP. Again, these two specific similar results are from two different universities 20 years apart.

2. Equity of drop rates within class types

Next, we can think of equity within each particular class type using an equity of parity model of equity. We will be using the idea of Gutiérrez [24], that demographic equity means that a student's success should not be predictable from their demographic characteristics. First, we have to say that inequity is easier to quantify than equity. For instance, in NonCLASP courses at NU, the data show us that 7.7% of students from historically marginalized groups dropped the course but only 4.7% of their peers dropped the course. That difference amounts to about three *extra* marginalized students, on average, dropping the course every year. If we assume that the NonCLASP courses were fully equitable and use a χ^2 test (the details of these χ^2 tests can be found in Appendix A) to measure how likely it is

that the extra drops by marginalized students are just a statistical fluctuation then we find that there is only a 2% chance (a p -value of 0.02) of a statistical fluctuation this large for a course that meets equity of parity. Most researchers would use values of p smaller than about 0.05 as evidence that this NonCLASP course is *not equitable* [31]. On the other hand, in the NU CLASP courses, there was less than one *extra* marginalized student per year dropping the course and a χ^2 test gives us $p = 0.31$ so in this case, on the assumption that CLASP courses are equitable, we have a 31% chance that this result is a statistical fluctuation. For this p -value, we would say that the data do not provide enough evidence to conclude that the course is inequitable (which is not quite the same as concluding that the course is fully equitable). In Appendix A, we use χ^2 tests for each of the demographic groups in Fig. 1 and their peers to quantify what we can already see fairly clearly from this figure itself. Specifically, this figure shows that the groups in the CLASP courses have drop fractions all roughly within a standard error of the overall average. Conversely, in the NonCLASP courses, some groups have drop rates considerably farther from the average. Even though we cannot prove the CLASP courses are equitable, we do use these results to suggest that CLASP is more equitable than NonCLASP, at least in terms of who drops the course.

VII. PASSING THE COURSE

The second step in succeeding in the first course of these intro-physics series (after not dropping the course) is to receive a grade of C- or better so that the course need not to be repeated. So, the next comparisons we make between these courses are finding the fractions of students who succeeded by receiving C- or higher grades in the first course in the sequence. We also examine equity for this type of success.

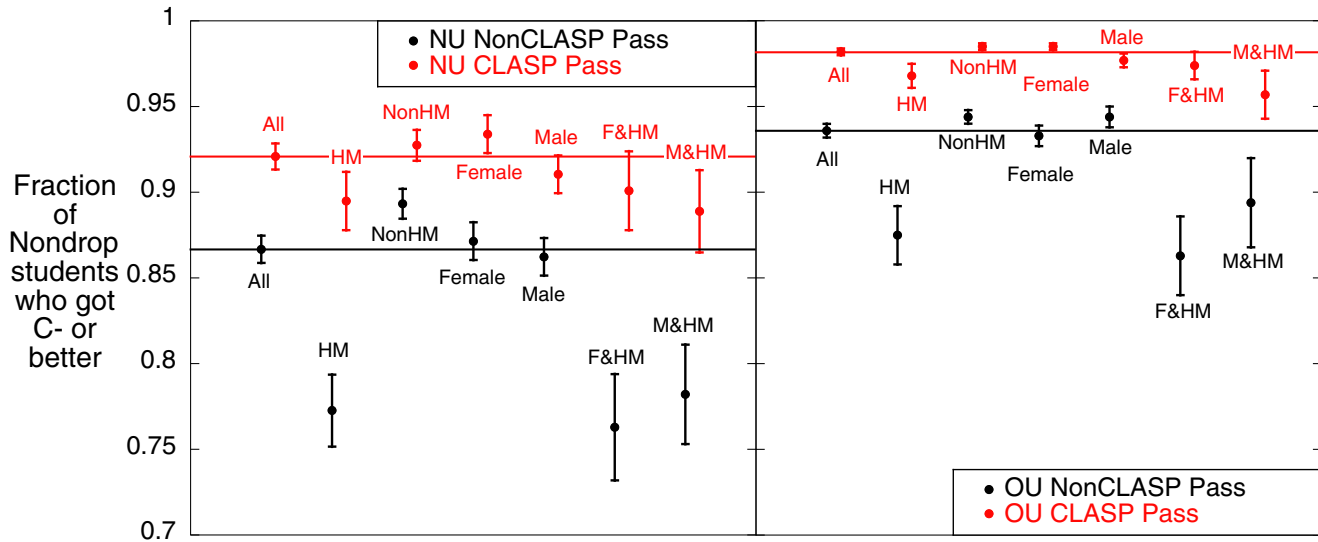


FIG. 2. The students who did not drop the course received grades. This plot shows the fraction of those students who received a grade of C- or better (in some important ways this amounts to passing the course). These grade fractions are shown for NU CLASP and NonCLASP classes on the left and for OU on the right. The error bars are standard errors. Again, several different ways of separating the students into demographic groups are shown.

A. Graphical examination of passing

Figure 2 shows the fraction of students that received a grade higher than D+ and therefore passed the course. The CLASP courses at both universities saw the highest level of success by this measure, indicated by the red lines being higher than the black lines in both cases. In addition to having these higher success levels overall, the CLASP courses appear to be more equitable than nonCLASP in the sense of the equity of parity model because the averages for each demographic group appear much closer to the average line for the students who took CLASP courses averages than they do for the students who took nonCLASP courses. Similarly, we can use an equity of individuality model and see that for each demographic group shown, the CLASP

course shows an improved pass rate, thus satisfying equity of individuality. (For further graphical breakdown of demographic groups by gender and race or ethnicity, see Figs. 7 and 8 in Appendix E.)

B. Statistical examination of passing the course

1. Pass -rate changes between NonCLASP and CLASP courses

In Table IV, we show that significantly more students are passing overall in the CLASP courses. At Next University, we see an effect size of about 21 more students per year passing, and at Original University we see about 53 more

TABLE IV. Class-type comparisons of passing the course for specific groups. For each university and for each demographic group in Fig. 2, we list the number of students in the measurement, the odds ratio for passing the course = (odds of passing a CLASP course)/(odds of passing a NonCLASP course), the standard error for those odds ratios, the *p*-value for those odds ratios, and the estimated net number of students *per year* from that demographic group who would not have passed the NonCLASP course but would benefit from the CLASP course because they passed their class. The standard error on this number of students is in parentheses.

Group	NU				OU			
	<i>N</i>	Odds ratio (SE)	<i>p</i> -value	Extra CLASP passes/year (SE)	<i>N</i>	Odds ratio (SE)	<i>p</i> -value	Extra CLASP passes/year (SE)
All	3095	1.79 (0.22)	<10 ⁻³	21.1 (4.2)	8004	3.75 (0.49)	<10 ⁻³	52.6 (5.4)
HM	728	2.50 (0.54)	<10 ⁻³	11.1 (2.5)	950	4.3 (1.2)	<10 ⁻³	12.5 (2.6)
NonHM	2107	1.53 (0.25)	<10 ⁻³	9.0 (3.3)	6733	3.84 (0.60)	<10 ⁻³	39.0 (4.6)
Female	1445	2.07 (0.41)	<10 ⁻³	11.2 (2.8)	4736	4.79 (0.86)	<10 ⁻³	35.6 (4.3)
Male	1650	1.63 (0.26)	0.003	9.9 (3.2)	3216	2.54 (0.50)	<10 ⁻³	15.2 (3.2)
F and HM	355	2.82 (0.88)	0.001	6.1 (1.7)	597	5.9 (2.2)	<10 ⁻³	9.5 (2.1)
M and HM	373	2.23 (0.66)	0.007	5.0 (1.8)	353	2.6 (1.1)	<10 ⁻³	3.2 (1.5)

students passing per year. The impact is similar given that Original University enrolls more students per year.

We can use the equity of individuality model to examine the impact on each demographic group. We use Eq. (2) but with DropOdds replaced with PassOdds and fit each demographic group in Fig. 2 including all of the students who did not drop and so received a grade. The results for odds ratio = (odds of passing CLASP with grade of C- or better)/(odds of passing NonCLASP with grade of C- or better) and standard error are given in Table IV. This table also gives the estimated net number of students per year (along with the standard error) who would have failed a NonCLASP course but passed a CLASP course. As expected from this figure, all of these odds ratios are larger than one. The largest odds ratio across both universities is for female students from historically marginalized groups who had about 3 times higher odds of passing CLASP at NU and about 6 times higher odds of passing CLASP at OU. Again, similar results at very different universities.

2. Equity of pass rates within class types

We have already noted the difficulty in using quantitative methods to prove equity of parity. For this reason, our quantitative analyses regarding equity of parity will be presented in Appendix A. However, we also note that this type of equity is already visible to the eye in Fig. 2, where we see that only for the CLASP courses does each demographic group have pass rates near the overall average so, although we may not argue that CLASP is fully equitable, it does appear that CLASP courses are more equitable than NonCLASP.

VIII. CLASP STUDENTS PREPARED FOR LATER STUDY

The algebra-based intro-physics series at NU consists of two semesters but only the first semester, PhysicsA, was changed to the CLASP format. The notably low drop rates, higher pass rates, and CLASP curriculum itself, which covers somewhat different content and emphasizes different skills than the traditional course it replaced, leaves open the possibility that students may be leaving the CLASP course insufficiently prepared for the following PhysicsB non-CLASP course and, indeed, for the rest of their work in their fields. We can quantify the large demographic effects on PhysicsB that may be due to CLASP PhysicsA's low drop and fail rates. The CLASP PhysicsA courses fed $68\% \pm 2\%$ of their nonHM students into PhysicsB and $67\% \pm 3\%$ of their HM students. In contrast, the NonCLASP PhysicsA courses that were supplanted sent $67\% \pm 1\%$ of their nonHM students but only $55\% \pm 2\%$ of their HM students. So the inequities of the NonCLASP courses and the equity of the CLASP courses are obvious in this measure. It is also clear that the CLASP PhysicsA courses changed the demographics of the PhysicsB

courses. First, we will look at average grades in PhysicsB at NU and then at the upper-division GPAs of these students in order to find out if there is a noticeable effect of sending some 22% more HM students into PhysicsB and beyond.

Because of the separation in time, between the NonCLASP PhysicsA courses, 2010–2014, and the CLASP PhysicsA courses, 2014–2017, these students took different PhysicsB courses. So, first we wanted to test whether there were any noticeable grading changes in the PhysicsB courses from the period 2010–2014 to the period 2014–2017. Fortunately, these courses all include a group of students who have never enrolled in any PhysicsA course, because they took a course equivalent to PhysicsA somewhere else. This group of students can be used, as a kind of standard, to compare the PhysicsB courses across time. Over 2010–2014, these students had an average PhysicsB grade of 2.71 ± 0.06 and over 2014–2017, they had an average grade of 2.68 ± 0.06 so it appears that there is no measurable change in the grading over these years. The NonCLASP PhysicsA NonHM students had 2010–2014 an average PhysicsB grade of 2.65 ± 0.03 and the HM students had an average grade of 2.51 ± 0.07 . For the CLASP PhysicsA students over 2014–2017, those numbers were 2.57 ± 0.04 for NonHM students and 2.47 ± 0.07 for HM students. Figure 3 shows these results. We suggest that these small effects (Cohen's d of 0.08 ± 0.06 for NonHM and 0.03 ± 0.17 for HM) between the CLASP and NonCLASP students in PhysicsB, even if they are not simply statistical fluctuations, show that the CLASP program is not only more equitable but also provides enough background for the students to succeed in their next physics

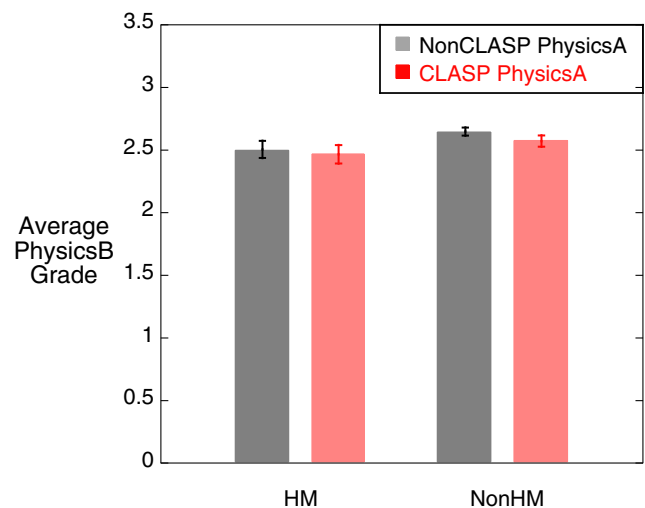


FIG. 3. The CLASP versions of the PhysicsA course sent over 20% more HM students into PhysicsB than the NonCLASP versions of PhysicsA but the entire set of CLASP students appeared to be as prepared as the NonCLASP students to succeed at PhysicsB, which was never changed to the CLASP format.

course. One point we want to emphasize here is that despite the fact that CLASP A at Next University sent 22% more HM students to Physics B, this did not result in a statistically different outcome grade wise for HM students.

Next, we examine upper-division GPA's for these students to compare CLASP PhysicsA students with NonCLASP PhysicsA students. We want to compare students taking roughly the same upper-division major courses so we will narrow down the group to students who majored in the 19 subjects that make up almost 90% of the majors taking the course. The graduation years 2017–2018 had about the same number of CLASP and NonCLASP students graduating so we will examine the GPAs for the final 2.5 years of these students' careers and take these GPAs as a proxy for their upper-division major GPA. We only include students who had at least 4 terms of study after taking PhysicsA, who had at least 43 units in that time, who did not transfer more than 20% of their units during this time, and we remove any PhysicsA grade points and units from the GPA. Our resulting sample size is $N = 487$. We find that the NonCLASP students graduating in those 2 years had an average upper-division GPA of 3.02 ± 0.03 and the CLASP students had an average GPA of 3.16 ± 0.03 . It might seem like the CLASP students thus did better in their later courses but, unfortunately, there is a confounding variable that seems to produce this apparent success. A look at the data shows us clearly that students graduating in these 2 years had higher GPAs if they took PhysicsA closer to graduation. This amounts to some sort of time-to-degree measure and turns out to hold for students from both the CLASP and NonCLASP courses. If we control for this time-to-degree-like variable, then the

difference in upper-division GPA between the CLASP and NonCLASP courses becomes 0.022 ± 0.043 so the GPA effect is removed and we conclude that the two groups are roughly equally prepared for upper-division work. This result does not change if we also control for students' academic ability by controlling for a student's GPA upon entering PhysicsA.

So, for both later physics work as well as later work in their major, it appears that the students who passed through CLASP courses were prepared as well as the students who passed through the NonCLASP courses. Of course, one might already think that one semester will not measurably change the academic trajectory of a group of students, however, we will see that by a different measure, STEM graduation rates, even one semester may have an effect.

IX. STAYING AND GRADUATING IN STEM

Previous work has suggested that grades in introductory STEM courses may well affect graduation rates, particularly for students from underrepresented groups [4]. The years included in our datasets are all long enough ago that we can expect to measure graduation rates for the students in our dataset so that we can compare STEM graduation rates for CLASP and NonCLASP courses. Again, we note that at the Original University an entire year of courses was changed to CLASP but that at the Next University only the first semester of the algebra-based introductory physics series was changed to the CLASP format.

For STEM graduation, we use the same years that we have been using previously and include all students who enrolled in the first term of the appropriate series and who

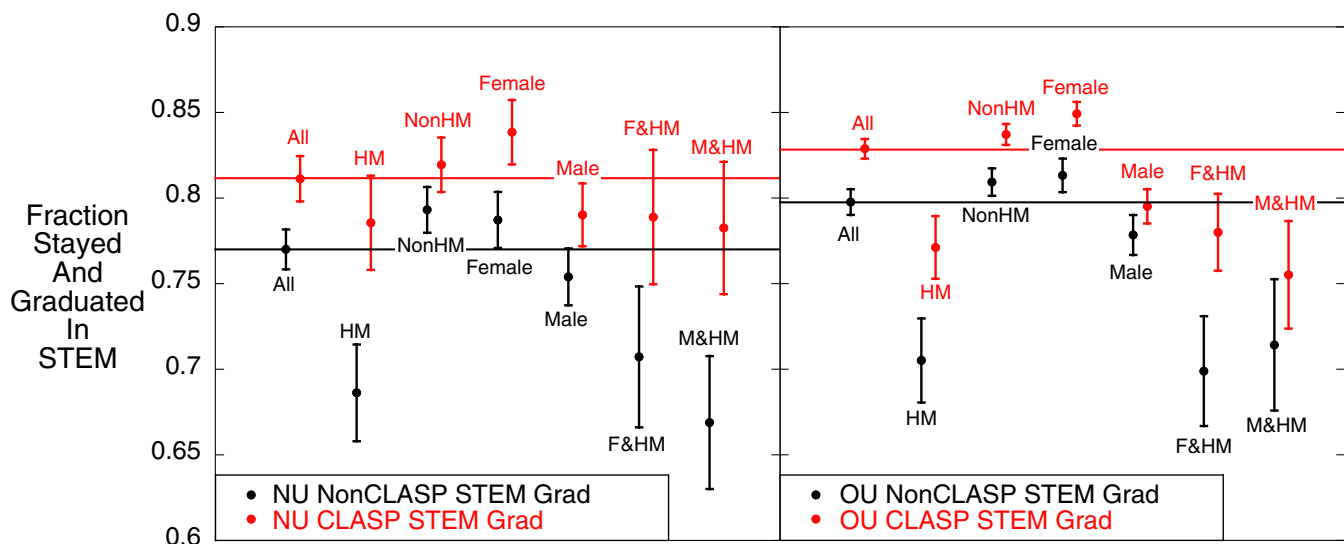


FIG. 4. Students who were STEM majors when they enrolled in the first term of intro-physics either graduated in a STEM field or did not. This plot shows the fraction of those students who graduated in STEM after starting out in STEM. These graduation fractions are shown for NU CLASP and NonCLASP classes on the left and for OU on the right. The error bars are standard errors. Again, several different ways of separating the students into demographic groups are shown.

had a STEM major upon that enrollment. We then find the fraction of those students who eventually graduated and who had at least one graduating major in a STEM field. These STEM graduation rates are shown in Fig. 4 for the same demographic groups that we have been using previously. We find that all demographic groups shown were at least as likely to graduate in STEM if they took a CLASP course than if they took a NonCLASP course with most groups being distinctly more likely to stay in STEM and graduate. Given what we already knew about the changes in STEM graduation at OU (increased graduation rates but not enough to approach equitable), we were surprised that STEM graduation after a CLASP course at NU is consistent with our definition of equitable, at least among the demographic groups shown. Even though students taking CLASP courses at OU graduated at higher rates, the distribution still seems far from equitable.

Although the data in Fig. 4 may be used to compare graduation rates across demographics in order to judge whether a course is equitable, the absolute values of these graduation rates are harder to analyze because the university-wide graduation rates at OU and at NU were each slowly increasing with time over the times in our database. This means that the increases that appear to happen when courses are changed to the CLASP structure may simply reflect a time-dependent increase in the overall graduate rates. For NU, we have graduation data on enough students across the university so that we can control for the time-dependent effects but we do not have detailed data on all of the students enrolled at OU during these times. We will limit ourselves to analyze the effects of CLASP at NU.

To better understand the higher graduation rates under CLASP at NU, we will construct a kind of difference-in-difference model (difference between CLASP and NonCLASP on top of possible differences across time) to allow us to separate a possible time-dependent background from our physics student data and minimize the effects of the time differences inherent to the changeover to CLASP. For this model, we chose a control group from the students at NU who did not ever take either the NonCLASP or the CLASP course and who were enrolled at some time in the years of our study, 2010–2017. We can use ethnicity and gender directly in our analyses so our control group just needs to have a large randomly chosen set of students with a variety of these characteristics. However, the graduation probability depends rather strongly on other student characteristics such as academic year and whether the student began at NU as a freshman or transferred from another school, so we would like a control group whose characteristics match those of our NonCLASP and CLASP courses. For these reasons, we identified a control group of 43,000 students chosen randomly but with weightings that gave us 15% freshmen, 33% sophomores, 26% juniors (12% transfers and 14% admitted earlier as Freshmen), and 26% seniors (14% transfers and 12% admitted as freshmen).

Of that group, we have ethnicity information for $N_{\text{control}} = 39,074$ control students. We added this control group to our database of $N_{\text{physics}} = 2938$ NonCLASP and CLASP students who were undergraduates and for whom we have ethnicity data. The control group includes both STEM majors and non-STEM majors so we will include both the STEM majors in the NU physics classes (68% of the students) as well as the NonSTEM majors and the graduation rates here will not be STEM graduation but simply graduation in any major.

To account for changes in graduation rates over time, we use a hierarchical logistic regression model. Specifically, the students will be grouped according to the academic term of their entry into NU. The hierarchical modeling fits each of these entry-term groups first and then assembles the resulting coefficients into overall final coefficients. We use the student's first NU term to group them because each of these groups will follow the same sets of university rules related to graduation [32]. Then we use the variable control , $= 1$ for the control group and $= 0$ for the physics students, to distinguish the physics students in our database from the rest of the university. We continue to use the variable CLASP , $= 0$ for students from NonCLASP classes and $= 1$ for students from CLASP classes, to measure the difference between these types of class. For each demographic group, we use this kind of hierarchical logistic regression method and fit the following model

$$\text{Log}(\text{GradOdds}) = b_0 + b_{\text{control}}\text{control} + b_{\text{CLASP}}\text{CLASP} \quad (4)$$

for the odds that a member of this group of students graduated. For each group, the fit produces the baseline overall graduation odds for NonCLASP students, the odds ratio comparing the control students to NonCLASP students, and the odds ratio comparing CLASP students to NonCLASP students. With these odds and odds ratio estimates, we can retrieve the average number of students per year who would have benefited from CLASP during these years.

The coefficients resulting from the regression fit to Eq. (4) are given in Appendix B in Table VIII. We use those odds and odds ratios, along with Eqs. (1) and (3), to find the graduation probabilities for each group in our model and show those probabilities in Table V along with their standard errors. Our main conclusions are seen in a comparison between CLASP and NonCLASP controlled graduation probabilities and in the estimated number of *additional* students graduating each year from NU who took the CLASP course. Even after controlling for university-wide changes in graduation rates, each demographic group in Fig. 4 has a higher graduation probability if they took CLASP than that group had after a NonCLASP class. The largest increase in controlled NU graduation rates came for female students from historically

TABLE V. Class-type comparisons for graduation for specific groups. The odds and odds ratios from fitting the logistic regression model, Eq. (4), are shown in Appendix B. We use those odds and odds ratios, along with Eqs. (1) and (3), to find the graduation probabilities for each group in our model. We show these probabilities below with their standard errors. The CLASP and NonCLASP graduation rates were then used to estimate the number of extra students graduating from CLASP each year who would not have graduated if they had taken the NonCLASP course. The control group comprises over 92% of N for each demographic group.

Group	N	Control		NonCLASP		CLASP		Extra CLASP Grads/year (SE)
		Graduation probabilities	Error	Graduation probabilities	Error	Graduation probabilities	Error	
All	46100	0.760	0.017	0.839	0.009	0.863	0.12	9.3 (6.9)
HM	13222	0.712	0.035	0.791	0.020	0.853	0.029	5.9 (3.4)
NonHM	28790	0.785	0.019	0.851	0.010	0.863	0.018	3.3 (5.7)
Female	22809	0.784	0.023	0.861	0.012	0.894	0.018	6.1 (4.1)
Male	23283	0.735	0.023	0.818	0.013	0.837	0.021	4.0 (5.3)
F and HM	7126	0.741	0.048	0.804	0.028	0.877	0.037	3.4 (2.2)
M and HM	6039	0.666	0.051	0.769	0.029	0.824	0.044	2.7 (2.6)

marginalized groups (examination of Fig. 4 suggests the same story would be true for OU). Of course, the result of these higher graduation rate estimates is that the extra number of CLASP students graduating is positive for each demographic group. The data in Table VIII can also be used to find the “graduation gap” for historically marginalized students compared to NonHM. This gap is estimated to be $7.3\% \pm 4\%$ for the control students, $6.0\% \pm 2.3\%$ for the NonCLASP students, but only $1.0\% \pm 3.4\%$ for the CLASP students. We consider this evidence that the CLASP course structure is more equitable than the NonCLASP structure and nearly closes the previously existing graduation gap.

Again, these results are for all marginalized students who took CLASP at NU during these years and not just for STEM majors and, also, the rate we are considering is for graduation in any major. In our entire dataset (which spans multiple years), there were 359 students from marginalized groups in CLASP classes so from a comparison of their CLASP graduation (in any major) rate of 0.853 with the same ethnicities’ NonCLASP graduation rate estimated to be 0.791, we see that 22 more students from marginalized groups graduated than one would predict if there had been no change to CLASP. Interestingly, if we use the numbers in Fig. 4 and do the same calculation of how many “extra” STEM majors from marginalized groups graduated with a STEM major we find that the number is 21. The fact that these numbers are almost the same suggests that the increase in the overall graduation rate that we found in our regression, for these particular students, is almost totally due to more CLASP students from marginalized groups choosing to stay in a STEM major and graduating with a STEM major.

X. DISCUSSION

This research is important for several different reasons. The first is that these kinds of studies are essential for progress to be made in higher education. While there is

almost universal agreement across fields of education that active learning environments are better for learning and other important student outcomes, there is still extreme skepticism among many faculty, (in part because implementing active learning often requires that some content is removed from the course) which prevents widespread reform from the department level up. Demands for these active learning reforms also rarely come from the top down perhaps because of the large initial cost of investment [33]. Therefore, one of the reasons research like this is useful is simply because it provides education research faculty an avenue to get peer-reviewed validation for their reform efforts to advocate for funding and convince their colleagues to continue teaching their courses using active learning. To be clear, research alone is not enough to change the culture around active learning, but it can be an essential step in the process.

Perhaps, the reason that is most important to the authors is that it provides further evidence for the productive use of a course deficit model. As discussed in detail in our previous publication [19], there is a persistent myth that equity gaps exist because some groups of students come in with lesser experiences and that the course structure is either irrelevant or inviolable. When one uses the student deficit model they generally ignore the possibility that the course controls these gaps. In using the student deficit model, students are often referred to as “not college ready” or “lacking preparation.” Our view is that the data support a course deficit model explanation. What a course deficit model does is reframe those equity gaps as a product of the course that is failing to leverage the assets of a diverse student population, rather than blaming the equity gaps on the students. With this approach, the system must adapt to the changing population because the alternative is to force our students to adapt to an outdated inequitable system. Because we are able to reduce and in some cases eliminate equity gaps simply by changing the structure of a single introductory course, this is a strong indication that the

university has much more power to eliminate equity gaps than they may realize.

This research provides additional evidence that introductory courses can potentially have a significant effect on student STEM retention, especially for marginalized students, and thus introductory courses are worth significantly investing in to boost long-term retention gains.

XI. CONCLUSIONS AND FUTURE DIRECTIONS

Our results indicate that students are more likely to pass CLASP introductory physics courses than the traditional courses they replaced with 4.9% more passing at Original University and 6.2% more passing at Next University. They also indicate that students are less likely to drop CLASP courses with 59% fewer students dropping at Original University and 81% fewer dropping at Next University.

We also show that historically marginalized students who take CLASP are more likely to graduate with a STEM degree. Taking CLASP at Next University is correlated with a 6% increase in graduation rate for HM students, and that is after controlling for the increase in graduation rates over time. This increase in historically marginalized student graduation rates effectively eliminates the HM student graduation gap for CLASP students.

The CLASP students at both institutions perform as well as the students who took traditional physics on later coursework, despite the fact that many more students are passing CLASP than passed the traditional course. By using a model of Equity of Individuality [23] to frame our analysis, we claim that implementing CLASP is an equitable practice as it resulted in increased success for most of the demographics groups we studied and no significant decreases for any groups. In some cases, there is also a strong case to be made for meeting equity of parity, meaning that HM students have similar outcomes to NonHM students.

There are also potentially other outside influences that may contribute to this change, but we have done our best to investigate those we know about. While skepticism is warranted around the claim that a single course can be responsible for such a significant change in retention in the major, these findings are consistent with the findings of Hatfield *et al.* [4]. They found that a single grade lower than a C- reduces the probability of a student's graduation with a STEM degree by about 15%.

This paper tells only a partial story regarding equity and success in CLASP. Pass rates for students in CLASP have changed in recent years at both institutions, indicating that some important elements of CLASP are tentative, and caution must be used before making claims that CLASP curriculum alone is responsible for improved success in physics and retention in STEM. Nevertheless, the success of the CLASP curriculum for thousands of student over many years strongly supports the course deficit model of

demographic grade gaps that we have previously argued for [19,25] using different sets of course changes.

If we operate under the conclusion CLASP does improve student retention in STEM, we can currently make few claims about what aspects of the CLASP curriculum contribute to these outcomes. However, there are several potentially fruitful avenues to pursue to explain our results. For example, since CLASP students are actively engaged in model construction and application instead of passively listening in lecture, they have more opportunities to successfully talk about science and use scientific tools. Moreover, these opportunities happen in a social environment where students are communicating with instructors and peers. At Next University, there is specific attention played to community building with activities throughout the semester that promote learning more about their classmates. This CLASP feature aligns with what research says is important for self-efficacy [34,35], identity [36], and belonging [37] factors for persistence in STEM.

Future work will examine how these potential factors contribute to CLASP success. We will also examine how our results change with time, paying particular attention to grading practices in the CLASP and situate our findings in relevant race and identity theories [36].

ACKNOWLEDGMENTS

The authors would like to thank the Physics Faculty at SJSU who reviewed an early draft of this work. This work is supported in part by the National Science Foundation under Grant No. 1953760.

APPENDIX A: QUANTIFYING EQUITY OF PARITY

1. Dropping the class

As discussed in Sec. VI B 2, our view of the equity of parity model of equity is that it matches Gutiérrez's [24] that demographic equity means that a student's success should not be predictable from their demographic characteristics. Examination of Fig. 1 suggests that dropping a CLASP course has less dependence on demographics than dropping a NonCLASP course. In this section, we quantify this issue. Specifically, for each independent demographic group in Fig. 1, we will use a χ^2 test to decide whether membership in that group is related to dropping the course. The results of these tests, for both course types and at each university, are shown in Table VI. When we find evidence that a group membership variable might be independent of the drop fraction variable (i.e., when the p -value is greater than 0.05), we place the associated large p -value in bold. In addition to the χ^2 results, we also include, as an effect size, an estimate of the number of students, n , in the listed groups who *would not have dropped* if their group had the same drop fraction as their peers. Table VI shows that, for either university, the CLASP courses more consistently produced these large p -values, suggesting that, by this

TABLE VI. Group-peer dropping the class comparisons within specific class types. Results of χ^2 tests of whether membership in the particular demographic group is statistically related to dropping the course. Each test has one degree of freedom and the total number of students in the test, N is shown as well as χ^2 and the p -value. Our definition of equity is when membership in a demographic group is not related to dropping the course so those tests with large p -values ($p > 0.05$ noted in the table by bold numbers) denote groups that dropped the course about as often as their peers. Also shown are effect sizes: n = estimated number of extra students per year dropping the course (as determined from the group-nongroup inequality in drop fractions). The standard error in n is in parentheses. The results shown in the left half of the table are for the NonCLASP and CLASP courses at Next University (NU) and the right half for the courses at Original University (OU).

Group	NU								OU							
	NonCLASP				CLASP				NonCLASP				CLASP			
	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$
HM	1758	5.47	0.02	2.8(1.4)	1185	1.04	0.31	0.6(0.7)	3162	9.83	0.002	4.5(1.9)	4586	1.12	0.29	0.8(0.9)
Female	1927	0.35	0.55	1.2(2.0)	1292	2.11	0.15	1.6(1.1)	3256	2.42	0.12	7.9(4.7)	4740	1.24	0.27	-3.6(2.8)
F and HM	1758	0.74	0.39	0.7(0.9)	1185	1.27	0.26	0.4(0.5)	3162	12.7	$<10^{-3}$	4.2(1.7)	4586	3.57	0.06	1.1(0.8)
M and HM	1758	4.81	0.03	1.8(1.0)	1185	0.04	0.84	0.1(0.4)	3162	0.17	0.68	0.4(0.9)	4586	0.66	0.42	-0.4(0.4)

TABLE VII. Group-peer passing the class comparisons within specific class types. Results of χ^2 tests of whether membership in the particular demographic group is statistically related to passing the course with a grade of C- or higher. Each test has one degree of freedom and the total number of students in the test, N is shown as well as χ^2 and the p -value. Those tests with large p -values ($p > 0.05$ noted in the table by bold numbers) denote groups that passed the course about as often as their peers. Also shown are effect sizes: n = estimated number of extra students per year who would pass the course (as determined from the group-nongroup inequality in pass fractions) if they had the pass rates of their peer groups. Values of n which are small (i.e., within estimated errors of their peers) are indicative of group-peer equity. The standard error in n is in parentheses. The results shown in the left half of the table are for the NonCLASP and CLASP courses at Next University (NU) and the right half for the courses at Original University (OU).

Group	NU								OU							
	NonCLASP				CLASP				NonCLASP				CLASP			
	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$	N	χ^2	p	$n(\text{SE})$
HM	1665	37.6	$<10^{-3}$	11(2.1)	1173	3.42	0.06	3.0(1.7)	3034	26.1	$<10^{-3}$	9.3(2.4)	4510	9.12	0.003	2.3(1.0)
Female	1821	0.19	0.66	-1.6(2.9)	1278	2.31	0.13	-4.2(2.7)	3125	1.77	0.18	7.8(5.8)	4663	3.28	0.07	5.3(2.8)
F and HM	1665	20.2	$<10^{-3}$	5.1(1.4)	1173	0.76	0.38	0.9(1.1)	3034	21.9	$<10^{-3}$	6.9(2.0)	4510	2.04	0.15	0.9(0.7)
M and HM	1665	12.7	$<10^{-3}$	4.4(1.4)	1173	2.28	0.13	1.6(1.2)	3034	4.41	0.04	2.2(1.3)	4510	8.75	0.003	1.3(0.7)

particular measure, they were more generally equitable than the NonCLASP courses that they replaced. These statistical results are consistent with our view of Fig. 1 both overall, in showing CLASP courses seem more equitable, and in the details, for instance showing that the NonCLASP courses led to some groups of marginalized students being much more likely to drop than their peers.

2. Passing the class

Again, we use the model of equity of parity, this time to examine the data displayed in Fig. 2 and, again, argue that the CLASP courses appear to be the most equitable. We quantify this assertion in the same way we did for data on dropping the course. We use χ^2 tests with each independent demographic group compared to their peers (the rest of the students). The results of these tests are shown in Table VII.

When the p -values associated with these tests are large (bold in the table), we argue that the pass rate of the group is relatively close to that of their peers and we compare these results between CLASP and NonCLASP within NU or within OU. Clearly, for CLASP courses, each of these demographic groups has pass rates close enough to their peers to yield large p -value results. In contrast, several of the NonCLASP courses have rather smaller p -value in these tests. Again, we do not use these results to argue that CLASP is fully equitable but just to argue that it appears more equitable than NonCLASP courses.

APPENDIX B: GRADUATION REGRESSIONS EQUITY OF INDIVIDUALITY

The odds ratios that were used to find the graduation probabilities in Table V and resulted from fitting the data to

TABLE VIII. Class-type comparisons for graduation, for specific groups. Shown are the odds (from b_0), odds ratios (from b_{control} and b_{CLASP}), errors, and associated p -values resulting from our fit of the logistic regression model, Eq. (4) and Eq. (3), to our data. The control group comprises over 92% of N for each demographic group. One can use the graduation odds for NonCLASP students together with the odds ratios for either the control students or for the CLASP students to find the graduation odds (and then the graduation rates) for those groups, respectively. The CLASP graduation rates were used to estimate the number of extra students graduating from CLASP each year who would not have graduated if they had taken the NonCLASP course. The standard errors are given in parentheses.

Group	N	b_0		b_{control}		b_{CLASP}		Extra CLASP grad/year (SE)
		NonCLASPOdds (SE)	p -value	OddsRatio (SE)	p -value	OddsRatio (SE)	p -value	
All	46100	5.23 (0.35)	$<10^{-3}$	0.60 (0.04)	$<10^{-3}$	1.20 (0.12)	0.073	9.3 (6.9)
HM	13222	3.79 (0.47)	$<10^{-3}$	0.65 (0.08)	$<10^{-3}$	1.53 (0.30)	0.026	5.9 (3.4)
NonHM	28790	5.73 (0.47)	$<10^{-3}$	0.64 (0.05)	$<10^{-3}$	1.10 (0.14)	0.444	3.3 (5.7)
Female	22809	6.20 (0.60)	$<10^{-3}$	0.59 (0.06)	$<10^{-3}$	1.35 (0.22)	0.067	6.1 (4.1)
Male	23283	4.50 (0.39)	$<10^{-3}$	0.62 (0.05)	$<10^{-3}$	1.14 (0.15)	0.314	4.0 (5.3)
F and HM	7126	4.11 (0.73)	$<10^{-3}$	0.70 (0.12)	0.039	1.73 (0.51)	0.060	3.4 (2.2)
M and HM	6039	3.34 (0.54)	$<10^{-3}$	0.60 (0.10)	0.002	1.40 (0.36)	0.189	2.7 (2.6)

Eq. (4) are given in Table VIII. Our focus is on the CLASP odds ratios derived from b_{CLASP} because those are (Odds graduating after taking CLASP)/(Odds graduating after taking NonCLASP). As expected all of these odds ratios are larger than one so even after controlling for possible time dependence in overall graduation rates, we still find that all of the demographic groups that we are considering graduated more often than the same groups taking NonCLASP classes. The same extra numbers of CLASP students per year graduating as in Table V are also shown in Table VIII.

APPENDIX C: ETHNICITY LISTS FROM OUR OU AND NU DATASETS

As pointed out in the main text, the institutions ask each student to identify their ethnic background but each institution limits the data that we have access to by recording only one result for the category of ethnicity (although at NU there is a category for two or more ethnicities that does not specify any particular ethnicity). In addition, the two institutions gather ethnicities in slightly

TABLE IX. The ethnicities as described in the data obtained from Original University.

Symbol	Ethnicity
AF	African-American/Black
CH	Chinese-American/Chinese
EI	East Indian/Pakistani
FP	Filipino/Filipino-American
IA	Indigenous American/American Indian/Native American
JA	Japanese-American/Japanese
KO	Korean-American/Korean
LA	Latino/other Spanish
MX	Mexican-American/Mexican/ Chicano
OA	Other Asian
WH	White/Caucasian

different ways. In our analyses in the main text we have combined students from marginalized ethnicities into a single group called HM but we wanted to show as much of the ungrouped data as made statistical sense to show. In this appendix, we will give a list of ethnicities included in our study as well as the label used for that group and, for groups making up at least 1% of the total [38], the number of students present in the dataset.

First, Table IX identifies the symbols we use for the ethnicities at OU that are most populated in our dataset and make up 89% of that dataset. About 7% of the students' ethnicities are unknown and there are four additional ethnicities represented in this OU dataset whose populations are too small, in either the NonCLASP classes or the CLASP classes or both, to be statistically useful.

Next, Table X identifies the ethnicities at NU that are most populated in our dataset and make up 81% of that NU dataset. About 9% of the students' ethnicities are unknown and about 10% are members of the 45 additional listed ethnicities at NU (most including fewer than ten students) represented in this NU dataset. At this point, we should note that we have previously shown [25] that at least one of these ethnicities, Asian, contains many groups that do not all seem to experience the same biases. Nevertheless, this grouping is already done for us in the NU dataset.

TABLE X. The ethnicities as described in the data obtained from Next University.

Symbol	Ethnicity
Asian	Asian
Blackprf	Black/African-American Preference
Filipino	Filipino
Hispa	Hispanic/Latino
White	White

APPENDIX D: DROPPING THE COURSE: MORE COMPLETE ETHNICITY AND GENDER DATA

A gender and ethnicity and class-type breakdown of the fraction of students who drop the class is shown for OU in Fig. 5. We find that no ethnicity and gender had significantly lower drops under NonCLASP courses.

A gender and ethnicity and class-type breakdown of the fraction of students who drop the class is shown for NU in

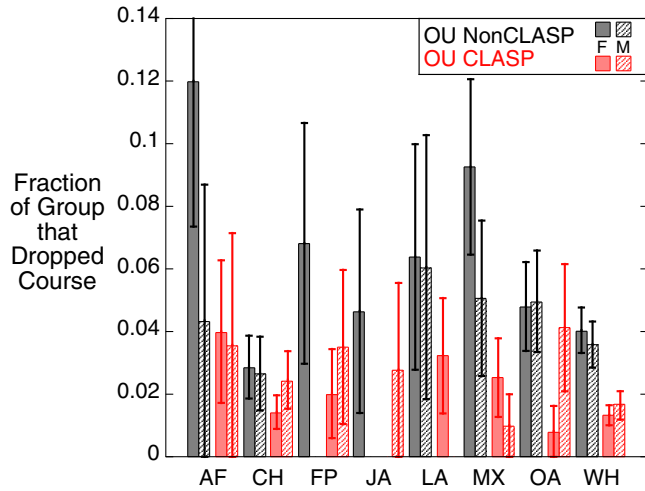


FIG. 5. Fraction of students dropping PhysicsA at OU separately showing (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols meanings are described in Table IX. If a group from one of the ethnicities seems missing, then that means that none of them dropped the course.

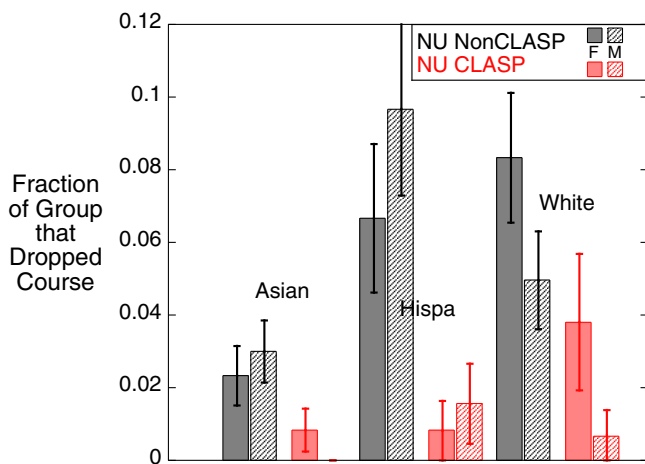


FIG. 6. Fraction of students dropping PhysicsA at NU separately showing (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols meanings are described in Table X. If a group from one of the ethnicities seems missing, then that means that none of them dropped the course.

Fig. 6. We find that no ethnicity and gender had significantly lower drops under NonCLASP courses.

APPENDIX E: PASSING THE COURSE: MORE COMPLETE ETHNICITY AND GENDER DATA

A gender and ethnicity and class-type breakdown of the fraction of students at OU who did not drop PhysicsA and did receive a grade of C- or higher is shown in Fig. 7. We find that no ethnicity and gender

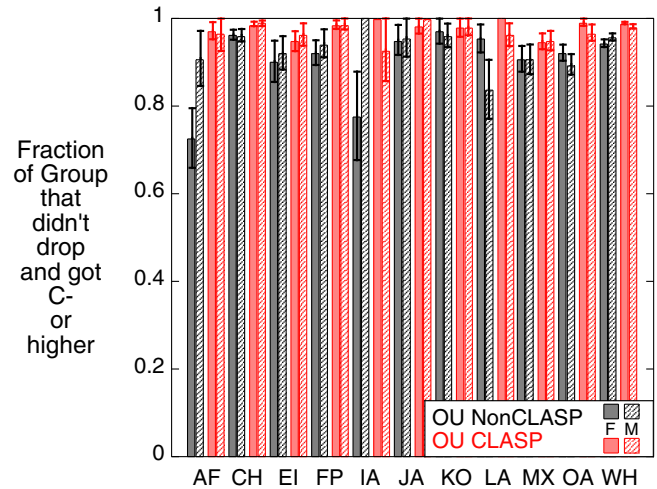


FIG. 7. Fraction of OU students who did not drop PhysicsA and did pass with a grade of C- or higher. These are separately shown for (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols' meanings are described in Table IX.

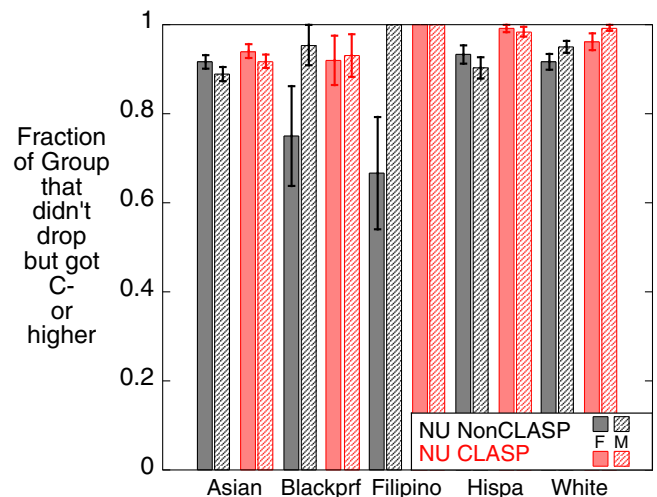


FIG. 8. Fraction of NU students who did not drop PhysicsA and did pass with a grade of C- or higher. These are separately shown for (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols' meanings are described in Table X.

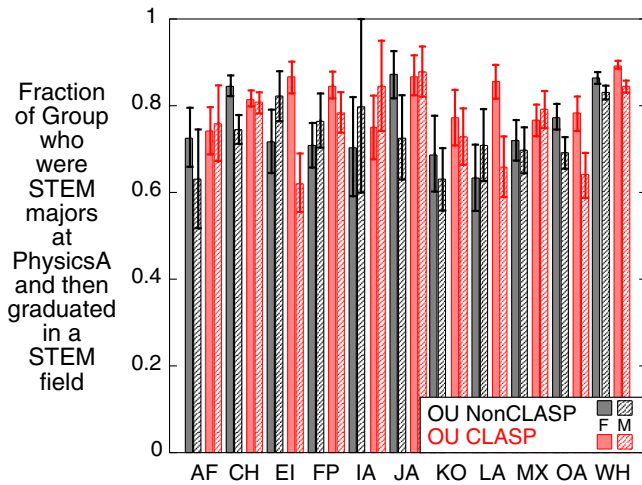


FIG. 9. Fraction of OU students who began PhysicsA as a STEM major and graduated with a STEM major. These are separately shown for (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols' meanings are described in Table IX.

had significantly higher pass rate under NonCLASP courses.

A gender and ethnicity and class-type breakdown of the fraction of students at NU who did not drop PhysicsA and did receive a grade of C- or higher is shown in Fig. 8. We find that no ethnicity and gender had significantly higher pass rate under NonCLASP courses.

APPENDIX F: STEM GRADUATION: MORE COMPLETE ETHNICITY AND GENDER DATA

A gender and ethnicity and class-type breakdown of the fraction of students at OU who began PhysicsA as a STEM

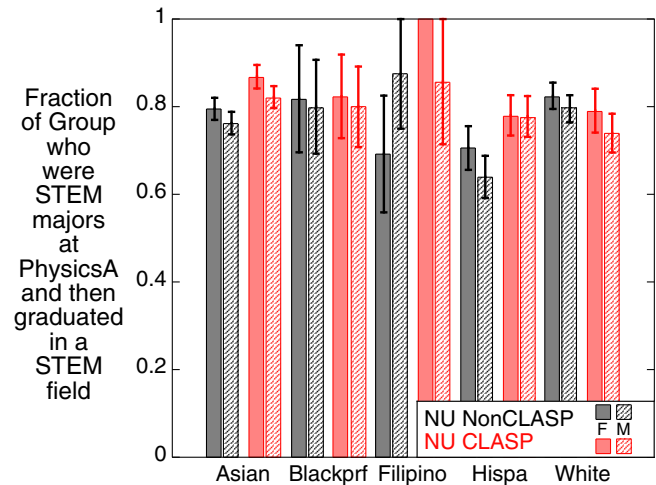


FIG. 10. Fraction of NU students who did not drop PhysicsA but did pass with a grade of C- or higher. These are separately shown for (i) the most populated ethnicities, (ii) each ethnicity separated by gender, and (iii) each ethnicity and gender group drop fraction by class type (CLASP or NonCLASP). The ethnicity symbols meanings are described in Table X.

major and graduated with a STEM major is shown in Fig. 9. Almost every ethnicity and gender graduated in STEM at a higher rate under the OU CLASP courses and none passed at a significantly lower rate.

A gender and ethnicity and class-type breakdown of the fraction of students at NU who began PhysicsA as a STEM major and graduated with a STEM major is shown in Fig. 10. Almost every ethnicity and gender graduated in STEM at a higher rate under the NU CLASP courses and none passed at a significantly lower rate.

-
- [1] A. Schuette, Navigating the enrollment cliff in higher education, Trellis Company, Technical Report, 2023, <https://files.eric.ed.gov/fulltext/ED628984.pdf>.
- [2] P. Copley and E. Douthett, The enrollment cliff, Mega-universities, COVID-19, and the changing landscape of U.S. Colleges, CPA J. (2020), <https://www.cpajournal.com/2020/10/05/the-enrollment-cliff-mega-universities-covid-19-and-the-changing-landscape-of-u-s-colleges/>.
- [3] N. Grawe, How to survive the enrollment bust, Chron. Higher Educ. **67** (2021), <https://www.chronicle.com/article/how-to-navigate-the-demographic-cliff>.
- [4] N. Hatfield, N. Brown, and C. M. Topaz, Do introductory courses disproportionately drive minoritized students out of stem pathways?, *PNAS Nexus* **1**, 1 (2022).
- [5] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410 (2014).
- [6] L. Kramer, E. Fuller, C. Watson, A. Castillo, P. D. Oliva, and G. Potvin, Establishing a new standard of care for calculus using trials with randomized student allocation, *Science* **381**, 995 (2023).
- [7] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [8] S. Kanim and X. C. Cid, Demographics of physics education research, *Phys. Rev. Phys. Educ. Res.* **16**, 020106 (2020).
- [9] W. Potter, D. Webb, C. Paul, E. West, M. Bowen, B. Weiss, L. Coleman, and C. De Leone, Sixteen years of collaborative learning through active sense-making in physics (CLASP) at UC Davis, *Am. J. Phys.* **82**, 153 (2014).

- [10] C. A. Paul, D. J. Webb, M. K. Chessey, and W. H. Potter, Equity of success in CLASP courses at UC Davis, presented at PER Conf. 2017, [10.1119/perc.2017.pr.068](https://doi.org/10.1119/perc.2017.pr.068).
- [11] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project, *Reviews in PER*, 2007, <http://www.per-central.org/document/ServeFile.cfm?ID=4517>.
- [12] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010106 (2010).
- [13] E. F. Redish, C. Bauer, K. L. Carleton, T. J. Cooke, M. Cooper, C. H. Crouch, B. W. Dreyfus, B. D. Geller, J. Giannini, J. S. Gouvea, M. W. Klymkowsky, W. Losert, K. Moore, J. Presson, V. Sawtelle, K. V. Thompson, C. Turpen, and R. K. P. Zia, NEXUS/Physics: An interdisciplinary repurposing of physics for biologists, *Am. J. Phys.* **82**, 368 (2014).
- [14] E. Etkina, D. T. Brookes, and G. Planinsic, *Investigative Science Learning Environment: When Learning Physics Mirrors Doing Physics* (Morgan & Claypool Publishers, Philadelphia, PA, 2019).
- [15] J. Von Korff, B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell, Secondary analysis of teaching methods in introductory physics: A 50 k-student study, *Am. J. Phys.* **84**, 969 (2016).
- [16] E. Etkina, K. Gibbons, B. L. Holton, and G. K. Horton, Lessons learned: A case study of an integrated way of teaching introductory physics to at-risk students at Rutgers University, *Am. J. Phys.* **67**, 810 (1999).
- [17] K. M. Cooper and S. E. Brownell, Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA students, *CBE Life Sci. Educ.* **15**, ar37 (2016).
- [18] D. Shafer, M. S. Mahmood, and T. Stelzer, Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both Asian American and African American students, *Phys. Rev. Phys. Educ. Res.* **17**, 010113 (2021).
- [19] D. J. Webb and C. A. Paul, Attributing equity gaps to course structure in introductory physics, *Phys. Rev. Phys. Educ. Res.* **19**, 020126 (2023).
- [20] S. L. Eddy and K. A. Hogan, Getting under the hood: How and for whom does increasing course structure work?, *CBE Life Sci. Educ.* **13**, 453 (2014).
- [21] S. Cotner and C. J. Ballen, Can mixed assessment methods make biology classes more equitable?, *PLoS One* **12**, e0189610 (2017).
- [22] R. R. Valencia, *The Evolution of Deficit Thinking: Educational Thought and Practice* (The Falmer Press, London, 1997).
- [23] I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer, Impact of equity models and statistical measures on interpretations of educational reform, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020103 (2012).
- [24] R. Gutiérrez, Context matters: How should we conceptualize equity in mathematics education?, in *Equity in Discourse for Mathematics Education*, Mathematics Education Library, edited by B. Herbel-Eisenmann, J. Choppin, D. Wagner, and D. Pimm (Springer, Germany, 2012), pp. 17–33.
- [25] C. A. Paul and D. J. Webb, Percent grade scale amplifies racial or ethnic inequities in introductory physics, *Phys. Rev. Phys. Educ. Res.* **18**, 020103 (2022).
- [26] B. Harrer and C. Paul, Modeling energy dynamics with the energy-interaction diagram, *Phys. Teach.* **57**, 462 (2019).
- [27] E. Kim and S.-J. Pak, Students do not overcome conceptual difficulties after solving 1000 traditional problems, *Am. J. Phys.* **70**, 759 (2002).
- [28] C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [29] W. Castillo and D. Gilborn, *How to "QuantCrit:" Practices and Questions for Education Data Researchers and Users*, EdWorkingPaper: 22-546 (EdWorkingPapers.com, 2023).
- [30] American association of physics website regarding diversity in physics, <https://www.aps.org/learning-center/statistics/diversity>, 2024 [accessed 25 July 2024].
- [31] Within the Course Deficit model the cause of the inequity is course structure. Hence the conclusion that the course is inequitable.
- [32] We would get nearly the same results and so draw the same conclusions if we grouped them by enrollment term but entry term is more appropriate for graduation.
- [33] E. Brewe, R. Dou, and R. Shand, Costs of success: Financial implications of implementation of active learning in introductory physics courses for students and administrators, *Phys. Rev. Phys. Educ. Res.* **14**, 010109 (2018).
- [34] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020111 (2012).
- [35] V. Sawtelle, E. Brewe, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *J. Res. Sci. Teach.* **49**, 1096 (2012).
- [36] H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, *J. Res. Sci. Teach.* **44**, 1187 (2007).
- [37] Y. Li, R. C. Bernardi, and E. Burkholder, The effects of active learning on students' sense of belonging and academic performance in introductory physics courses, *Eur. J. Phys.* **45**, 045705 (2024).
- [38] When the number of students is small then the statistics are useless and we also risk identifying a student.