

# Structuring groups for gender equitable equipment usage in labs

Matthew Dew<sup>1</sup>, Emily M. Stump<sup>1</sup>, and N. G. Holmes<sup>1\*</sup>

*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*



(Received 21 November 2024; accepted 3 June 2025; published 27 June 2025)

Previous research has found gender inequitable equipment usage across various lab course contexts. Few studies, however, have tested possible remediation strategies. In this work, we use hierarchical linear modeling to compare men and women's lab equipment usage in two group work structures across three course contexts. In one in-person course, students formed their own groups in class and rotated into new groups every unit. In the other two courses, one in person and one remote, students were assigned groups formed by the instructor and worked with the same group all semester. In line with former studies, we found gender inequitable equipment usage in the course with in-class formed, rotated groups. We did not observe gender inequitable equipment usage, however, in the course with instructor-assigned, fixed groups. Analyzing equipment usage across the semester within each course, our results suggest that this improvement comes from a combination of both instructor-assigned groups and keeping groups fixed for the semester. Our findings present many opportunities for subsequent controlled studies to probe these practices.

DOI: [10.1103/yyml-sr6q](https://doi.org/10.1103/yyml-sr6q)

## I. INTRODUCTION

Physics lab classes provide students with an opportunity to explore experimental physics practices that are not otherwise taught in lecture courses. These practices include tasks such as collecting data using equipment, analyzing data, and taking lab notes [1]. Introductory lab courses provide a structured environment for students to learn these skills early in their degree with regular feedback. Logistical constraints, however, can indirectly affect the access students have to learning these skills. For example, lab courses have limited equipment, which means students need to work in groups to collect data to complete their experiments. Group work is intended to give every student access to all parts of the lab and can also develop students' collaboration and communication skills, but it can also create inequities, which instructors do not intend. In this study, we evaluate the impacts of different instructional group work strategies at mitigating inequities in students' group work: namely in their equipment handling.

### A. Theoretical framework

In this work, we are interested in students having “equal access to the learning environment” [2]. Students can have

unequal access in labs in myriad ways, such as group members dominating conversation [3,4], ideas not being given proper consideration [3], or some group members having less access to certain tasks [5–10]. Here, we particularly focus on tasks related to handling the equipment, as access to the physical apparatus is a key part of the learning environment of instructional labs [1].

We assume that students' engagement with the physical apparatus impacts their learning. According to the “ICAP” framework, students' learning engagement ranges over four modes from most engaged to least engaged: interactive, constructive, active, and passive [11]. The ICAP framework argues that students learn more effectively with higher levels of engagement. For example, in the context of equipment usage, each of these levels of engagement could be mapped as follows:

- Passive: A student *receives* information about equipment by watching their group mate take a measurement.
- Active: A student *manipulates* equipment by taking measurements.
- Constructive: A student *generates* knowledge by devising their own equipment setup to collect data.
- Interactive: A student *and their group mate* generate knowledge via *dialoguing* by devising an equipment setup they could not come up with alone.

Lower equipment usage could mean a student is more likely to only be passively engaged with equipment and, therefore, learn less about handling, setting up, and devising experimental setups—important learning outcomes for introductory physics lab courses [1]. Thus, inequities in equipment use may in turn reflect inequities in learning.

\*Contact author: [ngholmes@cornell.edu](mailto:ngholmes@cornell.edu)

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

## B. Group work structures

Many studies of undergraduate introductory physics labs have found inequities in students' equipment usage. One study found men were more often equipment users than women in inquiry-based labs, but not in traditional labs [7]. Another study of traditional labs, however, found that men were more often their group's equipment user [9].

These inequities occur frequently enough that they have been named. In the “tinkerer and secretary” dynamic, men act as “tinkerers” who spend large amounts of class time using equipment while women act as “secretaries” who are responsible for their group's notes [9]. Such archetypes have been observed as early as grades 5–8 [12], while one study found them absent among students up to early secondary school [13].

Notably, these gendered differences are not due to student preferences. Two studies asked students in introductory physics labs what tasks they prefer to take on and these did not explain differences in task division [2,6]. These findings suggest that interventions to equalize access to roles should not attempt to differentially promote certain roles to students, but instead should change how the course is structured to overcome barriers to access.

Many such structures have been presented for mitigating inequities in group work, such as assigning students roles [7,14,15], having students share roles [16,17], avoiding isolating women in groups [14,18], and requiring explicit conversations about roles [2,7]. Most recommendations, however, are intended for lecture or recitation group work, rather than labs. Additionally, many recommendations are given without reference to empirical evidence [15,19–22].

One recent study, however, found that partner agreements were an effective intervention to mitigate gendered differences in equipment usage for an introductory mechanics lab [6]. In the lab courses in that study, newly formed groups completed a short written assignment that prompted them to discuss equitable distribution of lab roles. In the following weeks, groups were given follow-up assignments to reflect on their group work effectiveness. Students formed their own groups and changed groups every three class sessions.

The effectiveness of the partner agreements, however, was not universal across course contexts. While partner agreements remedied gendered equipment usage in the first-semester mechanics lab, they potentially introduced gendered differences in the electricity and magnetism lab course. This discrepant result could be due to lab practical final exams. The mechanics lab course had a lab practical final exam, which could have incentivized students to leverage partner agreements to gain experience with equipment. The electricity and magnetism lab course instead had a final group project lab so students were not individually tested on their equipment skills. Given that many physics labs do not have lab practical final exams as a grade incentive [23], it is worth exploring an

intervention that makes equipment usage more equitable across other contexts.

In a preliminary study, we found that our in-person lab course had inequitable equipment handling between men and women, but our remote lab course during the COVID-19 pandemic did not have these inequities [5]. In addition to being entirely remote, the groups were intentionally formed by the instructor (rather than formed by students in class, as defined Sec. II A) and constant for the semester (rather than changing every unit). We also found that groups typically took turns having one student make the experimental setup each week. The equitable participation, therefore, could be due to the remote modality, the new group structure, the emergency switch to remote learning, or some combination of these and/or other effects.

Other studies of equipment usage suggest that the group work structures may have supported equitable participation. One study with groups formed in class by students and rotated every period found no gender inequities in a traditional lab course, but found them in the inquiry-based version of the lab [7]. In the study with partner agreements, the control section for the mechanics lab—where students did not complete partner agreements, formed groups in class, and rotated groups—had gender inequitable equipment usage [6]. Another study of an inquiry-based lab found men used equipment more than women when groups were fixed for the semester, but how groups were formed was not specified [8]. A study of traditional physics labs found gender inequitable equipment usage and note taking with groups fixed for the semester; however, these groups were formed by students in class [9]. One study of student task division in inquiry-based labs had instructor-assigned pairs that rotated throughout the semester and found no inequities in equipment usage, though there were inequities in computer usage [10]. Collectively, these studies suggest that having instructor-assigned groups and keeping them fixed for the semester could remedy inequities in equipment usage.

Here, we test the effects of instructor-assigned, fixed groups as a plausible intervention for equitable equipment usage as compared to in-class formed, rotated groups. We postulate plausible mechanisms to further support carrying out this comparison. First, placing students in groups intentionally can create a good starting point for healthy group dynamics. Additionally, students placed in the same group for the entire semester, rather than changing every few weeks, have more opportunities to get to know their group mates and have a better understanding of what their group mates have done throughout the lab course. Because students will know when their group mates repeatedly do the same tasks, it is possible that students in semester-long groups are more likely to share equipment equitably across the semester (e.g., take turns). Frequently changing groups may not have this history and, thus, are more prone to inequitable access. With partner agreements, however,

students in frequently changing groups may have a structured means to negotiate for more equipment usage, but potentially only when they are given an incentive—such as a graded lab practical exam. While fixed groups could lead to students having the same role every week, our prior work suggests this is not the most common outcome [5].

### C. Current study

In this study, we have expanded upon this prior work of men and women's equipment usage by observing three courses: one in person with instructor-assigned groups fixed for the semester, one remote with in-class formed, rotated groups, and one in person with in-class formed, rotated groups. Notably, the course with instructor-assigned groups fixed for the semester also featured electricity and magnetism labs and no lab practical final exam—similar to the course in the prior study where partner agreements were potentially harmful [6]. We ask the following research questions:

- (1) Do lab groups that are instructor-assigned and fixed for the semester result in more gender equitable equipment usage within groups compared to groups that are in-class formed and rotated every unit?
- (2) If so, which is responsible for more equitable equipment usage: instructor-assigned groups, keeping groups fixed for the semester, or a combination of both?

Here, we define “equity” through the lens of participatory equity which “refers to the fair distribution of both participation opportunities and participation itself as students engage in the learning process” [24]. We operationalize this as men and women having the same level of equipment usage within their groups, on average, across the course.

## II. METHODS

There are three key components to our methods. In Sec. II A, we discuss the course context. In Sec. II B, we discuss the methodology for observations and the associated code book. In Sec. II C, we discuss the methods we used to analyze the results of our observations.

### A. Course context

We observed lab courses at Cornell University, an Ivy League university in upstate New York. These labs were skills-based and taught using the Structured Quantitative Inquiry framework [25]. This framework emphasizes developing students' critical thinking skills in a physics lab context. These lab courses were intended for students enrolled in the calculus-based introductory sequence, with primarily physics and engineering majors. The course mostly consists of first-year students. Details for the course-wide demographics can be found in Appendix A.

Students worked in groups of three or four to complete data collection and submit one set of collaborative lab notes as a group each week and were not provided directions about how to work effectively in a group. The teaching assistants were instructed not to allow groups of two or five but rarely these groups would form regardless. Students used lab notes to document their procedure, data analysis and calculations, and results. The lab notes were less formal than lab reports and were written entirely within each 2 h lab session. The final lab unit was a project lab, where each group had the opportunity to decide and investigate their own research question. Classes were taught by graduate student teaching assistants. Rarely, an undergraduate teaching assistant was also present.

We compared different group work conditions in the same physics introductory lab course across three semesters: one prepandemic, one mid-pandemic when classes were remote, and one postpandemic when classes returned to in-person instruction.<sup>1</sup> The primary difference in group work structure between each semester's lab course was how groups were set throughout each semester. For the prepandemic semester, students were not assigned groups—they formed groups every unit where each unit typically lasted two sessions. To check how often students switched group members, we calculated how many unique group members each student had over the semester. At minimum, a student could work with 1 unique group mate (if they worked in the same group of 2 all semester); at maximum, they could work with 12 unique group mates (if they grouped with 3 unique students all 4 units). For our dataset, students worked with  $6 \pm 2$  unique group members on average. This number is lower than the maximum because the average group size was about 3 and students usually changed members every unit, but occasionally they had one member in their group they had been grouped with prior.

For the mid-pandemic semester, students were in fixed groups for the entire semester to give them more stability in their courses during the COVID-era. Instructors also assigned students to their groups using results from a student preferences survey. Students completed their labs entirely remotely during this semester and did not share a lab space. The labs were very similar in content to previous semesters, albeit students made equipment out of materials they had around their homes (e.g., using a charging cord and block as a pendulum or using a bag of coins and a rubber band in place of a spring and weights). Students were given suggestions each lab for equipment alternatives

<sup>1</sup>The mid-pandemic observations were previously analyzed in Ref. [5]. We include them here because they provide an example of equitable equipment handling, but our analysis for the current study differs from the analysis in Ref. [5]. We fully recognize that we cannot disentangle the mechanisms for the observed equitable equipment handling in that semester due to the complex nature of the pandemic and the shift to emergency remote teaching.

with the intention of making equipment accessible for students in dorm rooms, at their family home, or somewhere else entirely.

Some groups chose to create multiple experimental setups, but most groups chose to only have one experimental setup per group each week. This meant typically only one student each week was responsible for equipment usage, which often prompted groups to explicitly discuss who would be taking measurements each week. These constraints may have inadvertently implemented some of the intentions behind partner agreements used in a prior study [6]; students were being prompted to discuss their roles without instructor intervention. These groups were also intentionally formed through the following procedure:

- (1) Students were grouped with classmates they requested (and not grouped with students they asked not to be grouped with) to give students some control and social comfort during the pandemic.
- (2) If there was any flexibility left, we avoided isolating women and students of a systemically nondominant race based on recommendations from prior research [14,18].

For the mid-pandemic semester, 13.4% of students listed at least one preferred group mate and 0.6% listed at least one person they preferred not to work with. The postpandemic semester was fully in person, and students were assigned fixed groups for the entire semester on nearly the same criteria as listed above in the remote course. Similar to the mid-pandemic semester, 14.2% of students listed at least one preferred group mate and 1.4% listed at least one student they preferred not to work with. Following recommendations from another study to attend to student preferences [2], we added one more criteria to our procedure in between steps 1 and 2 above:

- Students were placed in groups according to their preferred gender composition (majority women, majority men, or self-described).

For this item, 75.7% had no preference, 16.2% preferred a majority women group, 4.4% preferred majority men, and 3.7% self-described.

The lab courses had the same learning goals and similar activities, but were slightly different across the three semesters. In the prepandemic and mid-pandemic

semesters, the lab was attached to a four credit hours, introductory calculus-based mechanics lecture course. Nearly all students were co-enrolled in both the lecture and lab. Two students were only in the lecture for the prepandemic semester; five students were only in the lab for the prepandemic semester and one for the mid-pandemic semester. The prepandemic lab course had invention activities [26], which typically took 5 to 15 min at the start of lab. These activities introduced students to specific data analysis concepts and tools by having them construct their own mathematical equations to evaluate data. The mid-pandemic and postpandemic lab courses adapted these invention activities into prelab homework assignments using worked examples [27]. Additionally, the mid-pandemic course had students use makeshift home equipment to collect their data. For example, students often used coins and rubber bands in place of standardized weights and springs for a Hooke's law lab. The postpandemic semester lab was a separate, one credit hour course (and the lecture component was a three credit hour course), though students typically took the lab and mechanics course concurrently. The standalone lab course structure introduced a few changes to the course activities. First, the lab course added a 50 min lecture each week where students covered topics in experimental physics, such as data analysis and ethics. Additionally, the standalone lab course compressed the lab experiments of introductory mechanics and electricity and magnetism into one semester. This standalone course gained two weeks of lab sessions because it no longer skipped labs during the lecture's two midterm exam weeks. Two electricity and magnetism units with two sessions each were introduced. The original first session of lab was converted into a first-week homework assignment and one of the mechanics labs was changed from two sessions to one. A summary of the differences in the three course contexts is provided in Table I.

## B. Observations

For the prepandemic semester, we observed two class sections for students enrolled in introductory mechanics for engineering majors. For the mid-pandemic semester,

TABLE I. Summary of differences between each course from Sec. II A.

|                 | In person, in-class formed,<br>rotated groups | Remote, instructor-assigned,<br>fixed groups | In person, instructor-assigned,<br>fixed groups |
|-----------------|-----------------------------------------------|----------------------------------------------|-------------------------------------------------|
| Modality        | In person                                     | Remote                                       | In person                                       |
| Semester        | Prepandemic                                   | Mid-pandemic                                 | Postpandemic                                    |
| Group formation | In-class formed                               | Instructor-assigned                          | Instructor-assigned                             |
| Time with group | Rotated every unit                            | Fixed for semester                           | Fixed for semester                              |
| Course delivery | Attached to lecture course                    | Attached to lecture course                   | Separate from lecture course                    |
| Lab lecture     | None                                          | None                                         | 50 min once per week                            |
| Topics covered  | Mechanics                                     | Mechanics                                    | Mechanics and electricity and magnetism         |

we observed three class sections: one section for students enrolled in introductory mechanics for physics and astronomy majors and two sections for students enrolled in introductory mechanics for engineering majors. For the postpandemic semester, we observed four sections of the standalone lab course.

All observations were conducted via prerecorded video. For the mid-pandemic semester, students recorded their own lab sessions via Zoom recording. The in-person semesters (prepandemic and postpandemic) were recorded by the researchers using cameras in the corners of the room. For the in-person courses, the recorded sections were chosen based on which sections had the highest percentages of students that consented to being recorded for the research study. For the remote course, students were in Zoom rooms with only their group rather than an entire section, so we recorded every group where all students consented to participate. Audio was not utilized during coding and observations were based only on visual cues.

We used a coding methodology similar to that used in previous studies [5–7]. We coded student activity at 5 min increments as either equipment or other. While prior studies also coded for activities such as writing on paper or using a laptop or desktop computer [6,7], we restricted our study to equipment usage. In part, this decision was due to the consistently observed inequities in equipment usage but not other roles. In addition, students in these courses used computers both to submit digital lab notes and conduct data analysis, so coding for computer usage would not distinguish these roles. Finally, for the remote semester, students used their laptop cameras that allowed us to effectively make determinations of when students were using equipment, but made it untenable to determine when a student was performing a lab task on their computer (separate from the Zoom discussions).

We coded for equipment when the student was handling standard lab equipment or handling something that was clearly being used as equipment for their experiment. This includes activities such as holding a ruler to make measurements, collecting equipment for the experimental setup, or using their phone as a timer. We coded equipment usage for these scenarios as they encompass student activity in the “active” level of engagement in the ICAP framework. Because our research questions are focused primarily on participatory equity, we do not distinguish the other levels of engagement in this study. We also coded a student as using equipment when we could be reasonably certain they were doing that activity despite an obstructed view. For example, if a student stepped out of frame while using a piece of equipment during our time of interest, but there were other visual cues that indicated that student was giving measurements to one of their teammates. If we could not determine what a student was doing at a particular time stamp (e.g., they were clearly using their hands for some activity but something obstructed our view of their hands),

we checked up to 30 s before and after the time stamp to make a determination.

Observations were coded by two groups of researchers including graduate and undergraduate researchers. The mid-pandemic course data were coded first; this was done by three researchers, one of whom was one of the authors (M. D.). The prepandemic data were coded by this same author after sufficient interrater reliability was achieved with the mid-pandemic data. The postpandemic course data were coded by the same author and three new researchers. In both cases with more than one coder, we used video data from the corresponding semester as data for training and interrater reliability tests.

For the mid-pandemic course data, we obtained a Fleiss’  $\kappa$  [28] of 0.77. For the postpandemic course data, we obtained a Fleiss’  $\kappa$  of 0.88. A Fleiss’  $\kappa$  of 0.75 or higher indicates excellent agreement [29]. In both cases, we evaluated differences after coding; there were no discernible trends in disagreements between coders.

Hereafter, we refer to each course by the corresponding modality and group work structure. That is, we refer to the prepandemic course as the in-person course with in-class formed, rotated groups; the mid-pandemic course as the remote course with instructor-assigned, fixed groups; and the postpandemic course as the in-person course with instructor-assigned, fixed groups.

A summary of the demographics by semester can be found in Table II for students and group fraction data points and in Table III for group gender compositions. We obtained all gender data for the in person, in-class formed, rotated groups and remote, instructor-assigned, fixed groups courses from surveys where students self-reported their gender. For the in person, instructor-assigned, fixed groups course, student genders came from self-reported surveys; however, ten students did not complete surveys. We supplemented this data with university registrar data. While a binary composition of gender is an oversimplification (see [30] for further discussion), only a single student in our observational data identified as nonbinary, so we cannot make any quantitative claims about nonbinary students’ participation.

## C. Analysis

### 1. Hierarchical linear modeling

For our analysis, we were interested in how students share their group’s time on equipment. We measure this using the amount of time each student spends on equipment as a fraction of their group’s time on equipment within a lab session. We call this measure “equipment group fraction,” as in a previous study [6]. The equipment group fraction for student  $i$  in the  $j$ th class session,  $g_{ij}$ , is defined in Eq. (1),

$$g_{ij} = \frac{N_{ij}}{\sum_k N_{kj}}, \quad (1)$$

TABLE II. Gender demographics for the students and equipment group fraction data points across all three semesters. An equipment group fraction data point is a student's equipment group fraction from a single class session, so most students are within the dataset multiple times throughout the semester. The total does not necessarily equal the sum of men and women because some students identified as nonbinary or preferred to not share their gender.

|       | Number of students                         |                                           |                                              | Number of equipment group fraction data points |                                           |                                              |
|-------|--------------------------------------------|-------------------------------------------|----------------------------------------------|------------------------------------------------|-------------------------------------------|----------------------------------------------|
|       | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups | In person, in-class formed, rotated groups     | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
| Women | 18                                         | 14                                        | 51                                           | 126                                            | 101                                       | 429                                          |
| Men   | 18                                         | 10                                        | 38                                           | 122                                            | 63                                        | 325                                          |
| Total | 36                                         | 25                                        | 93                                           | 248                                            | 172                                       | 782                                          |

TABLE III. Group gender composition demographics for the observed sections and all sections across all three semesters. If the same group worked together  $n$  times in a semester, that appears  $n$  times in our counts below. A group's gender composition is "unknown" if at least one student preferred not to share their gender and the group was not definitively mixed gender.

|                        | Observed sections                          |                                           |                                              | All sections                               |                                           |                                              |
|------------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
|                        | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
| All women              | 20                                         | 9                                         | 61                                           | 153                                        | 50                                        | 179                                          |
| All men                | 19                                         | 0                                         | 45                                           | 104                                        | 74                                        | 139                                          |
| All nonbinary or other | 0                                          | 0                                         | 0                                            | 1                                          | 0                                         | 0                                            |
| Mixed gender           | 52                                         | 37                                        | 124                                          | 394                                        | 632                                       | 859                                          |
| Unknown                | 0                                          | 0                                         | 0                                            | 16                                         | 244                                       | 364                                          |

where  $N_{ij}$  is the number of observations of student  $i$  using equipment in class session  $j$ , and  $\sum_k N_{kj}$  is the total number of times we observed the members of their group using equipment in the  $j$ th class session.

As an example, consider a group of three students A, B, and C that we observe 10 times each over the course of a class session. Let us say we observe student A using equipment 2 times, student B using equipment 0 times, and student C using equipment 6 times. Student A would have an equipment group fraction of  $2/8 = 0.25$ , student B would have an equipment group fraction of 0, and student C would have an equipment group fraction of  $6/8 = 0.75$ .

For our data, each student is within a particular group anywhere from 1 to 9 times. Because students can be in the same group over multiple class sessions, we use hierarchical linear modeling. Hierarchical linear modeling, also known as multilevel modeling, is a regression method that accounts for random effects of data structured into levels (see Van Dusen and Nissen for a review of hierarchical linear modeling for physics education research [31]). In education research, hierarchical linear modeling most often uses data where individual students (level-1) are nested within classrooms (level-2). Each student is represented a single time in the dataset and there may be structural differences between students in different classrooms. Here,

individual students are nested within groups and may appear in the dataset multiple times. Thus, we define our level-1 data as a student's equipment usage in a single class session; that is, a single group fraction. We define our level-2 data as a particular student in a particular group (rather than a particular student across the entire semester), as in Ref. [6]. We chose this definition because groups are complex systems where students will likely behave similarly within the same group of students but we cannot adequately assume a student will behave the same across three different groups.

The level-1 data is modeled in Eq. (2). We model the group fraction for the  $j$ th session of student  $i$ ,  $g_{ij}$  such that

$$g_{ij} = \beta_{0j} + \beta_{1j} \text{GroupSize} + r_{ij}, \quad (2)$$

where  $\beta_{0j}$  is the intercept, GroupSize is the number of students in student  $i$ 's group for session  $j$ , and  $r_{ij}$  is the residual term. We include a variable for group size here because an equitable equipment group fraction decreases as group size increases. For a group of two students that share equipment equitably in a class session, each would have an equipment group fraction of 0.5. For a group of four students, each would have an equipment group fraction of 0.25.

The level-2 data is modeled in Eq. (3),

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{Gender} + u_{0j}, \quad (3)$$

where  $\gamma_{00}$  is the mean intercept, Gender indicates the student's gender with woman as the reference value, and  $u_{0j}$  is the residual term. We check that the assumptions of hierarchical linear modeling are met in Appendix B. Through our lens of participatory equity, a course with gender equitable equipment has a  $\gamma_{01}$  term that is statistically indistinguishable from 0 as that means, on average, men and women have similar equipment group fractions—or, in terms of the ICAP framework, men and women are equitably actively engaged with equipment—throughout the course. We ran separate models for each course condition rather than including an interaction term to improve interpretability.

## 2. Effects over time

To further differentiate the effects of fixed/rotated and instructor-assigned/in-class formed groups on men and women's equipment usage, we investigated the average equipment group fraction across class sessions. Each course has at least eight observed class sessions, but we do not have a sufficient number of students in each class session to make statistical comparisons. While a model that considers time evolution across the semester could work with our sample size, this will only tell us the general semester trend and not help us determine what aspects of each course structure are beneficial or detrimental. We are interested in how equipment group fraction varies as groups initially form (every unit in the rotated groups case or only once in the fixed groups case) and as the semester progresses.

Our measure of equipment group fraction, Eq. (1), however, is dependent on group size. While the earlier regression analysis accounted for group size, equipment group fraction alone does not. This presents difficulties comparing data across groups without regression. Recall our earlier example with students A, B, and C where the students have equipment group fractions of 0.25, 0, and 0.75, respectively. Now consider a group with students D and E who have equipment group fractions of 0.25 and 0.75, respectively. In a group of three students, an equipment group fraction of 0.25 is not as far from an equal share (one third) as it is in a group of two (one half). Thus, here we normalize each group fraction by the group size to account for different standards for equity across group sizes.

The normalized equipment group fraction for student  $i$  in the  $j$ th class session,  $g_{ij,\text{norm}}$ , is calculated according to Eq. (4),

$$g_{ij,\text{norm}} = g_{ij} N_{ij,\text{students}}, \quad (4)$$

where  $N_{ij,\text{students}}$  is the number of students in student  $i$ 's group in class session  $j$  and  $g_{ij}$  is the associated equipment group fraction. In a group of  $n$  students with equitable equipment sharing, each student would have an equipment group fraction of  $1/n$ ; therefore, a normalized group fraction of 1 is considered equitable. This provides a metric that shows how much more ( $> 1$ ) or less ( $< 1$ ) a student used equipment than would be equitable in a group of any size.

Returning to our previous examples, student A's equipment group fraction of 0.25 would be normalized to  $0.25 * 3 = 0.75$ , student B's equipment group fraction of 0 would remain  $0 * 3 = 0$ , and student C's equipment group fraction of 0.75 would be normalized to  $0.75 * 3 = 2.25$ . For the other group, student D's equipment group fraction would be normalized to  $0.25 * 2 = 0.5$  and student E's to  $0.75 * 2 = 1.5$ . Now we can more appropriately compare students A and D's equipment usage (as well as students C and E) in terms of closeness to an equitable share.

## 3. Supplementary analysis

Equipment group fraction is an effective measure for answering our research questions regarding equity within groups; however, it could dilute the effects of groups with varying levels of total equipment usage. For example, a student with an equipment group fraction of 0.25 may be in a group that used equipment throughout the whole lab, or they may be in a group that barely touched the equipment. In Appendix C, we consider a different metric of student equipment usage, equipment individual fraction, that instead focuses on how individual students allocate their time in lab. This alternate measure provides additional context for equitable equipment usage, but is limited in its ability to interpret equitable shares of equipment handling within groups. This analysis acts as alternative supporting evidence for our main argument. We find similar trends to those using equipment group fraction.

## III. RESULTS

The results from the equipment group fraction regression are shown in Fig. 1 and Table IV. For the in-person course with in-class formed, rotated groups, we found that men were responsible for more equipment usage than women within their groups ( $\gamma_{01} = 0.118 \pm 0.042$ ,  $p = 0.005$ ) controlling for group size and random effects. We do not observe a statistically significant gendered difference for the remote course with instructor-assigned, fixed groups ( $\gamma_{01} = -0.109 \pm 0.077$ ,  $p = 0.159$ ) or the in-person course with instructor-assigned, fixed groups ( $\gamma_{01} = 0.014 \pm 0.030$ ,  $p = 0.640$ ).

The results separated by class session for each course are shown in Fig. 2. For the in-person course with in-class formed, rotated groups, men were responsible for more of

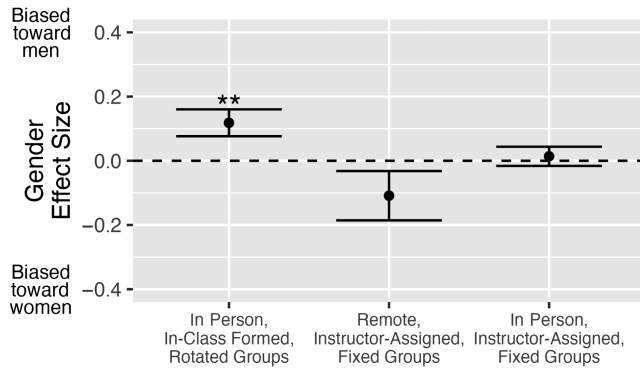


FIG. 1. Results from multilevel regression for equipment group fraction controlling for group size and random effects. A value greater than zero indicates men were responsible for more equipment usage in their groups than women, whereas a value less than zero indicates less. The error bars represent the standard error of the regression coefficients. The asterisks denote statistical significance where \*\* indicates  $p < 0.01$ . The full model output can be found in Table IV.

their group's equipment usage at the start of the semester; this gap decreased as the semester progressed, but men were consistently responsible for more equipment usage. We also notice possible small differences between the first and second sessions of each unit, where gaps between men's and women's equipment usage are larger in the first unit of each session, when new groups were formed (particularly unit 1 session 1 [U1S1] vs unit 1 session 2 [U1S2], U2S1 vs U2S2, and—to a less extent—U3S1 vs U3S2). For the remote course with instructor-assigned, fixed groups, neither men nor women were consistently responsible for more equipment usage; however, our uncertainty is very large due to a small number of data points. For the in-person course with instructor-assigned, fixed groups, there was no difference between men and women's equipment usage at the start of the semester. Additionally, there was no consistent trend in differences in equipment group fraction across the semester.

#### IV. DISCUSSION

In this study, we compared the equipment usage of men and women within their groups across three different lab

course contexts. In one of these courses, offered in person, students formed groups in class and rotated groups every unit. In the other two courses, one offered remotely and one in person, students were assigned to groups intentionally by the instructor and remained with their groups all semester.

Our results show men used equipment more than women, on average, in the in-person course using in-class formed, rotated groups, much in line with prior studies with similar group work structures [6,7]. For the courses with instructor-assigned groups fixed all semester, however, men were not responsible for more equipment usage than women within their groups. Expanding upon our former study [5], this result tentatively suggests that instructor-assigned groups fixed for the semester improved the equity of equipment usage. The three conditions, however, include (at least) three different instructional variables that may have caused this result.

First is the role of in person vs remote instruction. Comparing the two courses with instructor-assigned, fixed groups, we find that men do not use equipment more than women in either course. In fact, women were responsible for slightly *more* of their group's equipment usage than men in the remote course. Possibly due to our remote course's small sample size, the difference between group fraction in these two courses is not statistically distinguishable and, therefore, is an area for further study.

Second is the role of fixed vs rotated groups. Comparing the two in-person courses, we find a gender difference only in the course with rotated groups, whereas men and women were responsible for similar equipment usage within their groups in the course with fixed groups. These results disagree with some studies that found gender inequitable equipment usage with groups fixed for the semester [8,9]. These studies and ours, however, not only modified whether or not groups changed, they also varied in how groups were formed in the first place.

Thus, our third variable is the role of in-class formed and instructor-assigned groups. Between the two in-person courses, only the course with instructor-assigned groups did not have a gender difference in equipment group fraction. Many of the contexts for the studies previously cited with gender inequities had groups formed in class by students [6,7], which is the condition in our study with the largest gender inequities. Together, our data and the previous studies

TABLE IV. Results from linear regression for equipment group fraction. The table includes the regression coefficients, standard errors, and  $p$  values in parentheses. The conditional (marginal)  $R^2$  values for these models are 0.585 (0.101) for in person, in-class formed, rotated groups; 0.169 (0.026) for remote, instructor-assigned, fixed groups; and 0.380 (0.040) for in person, instructor-assigned, fixed groups.

| Predictor           | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
|---------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
| Intercept           | $0.719 \pm 0.130$ ( $<0.001$ )             | $0.696 \pm 0.328$ (0.035)                 | $0.650 \pm 0.088$ ( $<0.001$ )               |
| Group size          | $-0.146 \pm 0.045$ (0.001)                 | $-0.100 \pm 0.085$ (0.243)                | $-0.098 \pm 0.024$ ( $<0.001$ )              |
| Gender (ref: women) |                                            |                                           |                                              |
| Men                 | $0.118 \pm 0.042$ (0.005)                  | $-0.109 \pm 0.077$ (0.159)                | $0.014 \pm 0.030$ (0.640)                    |

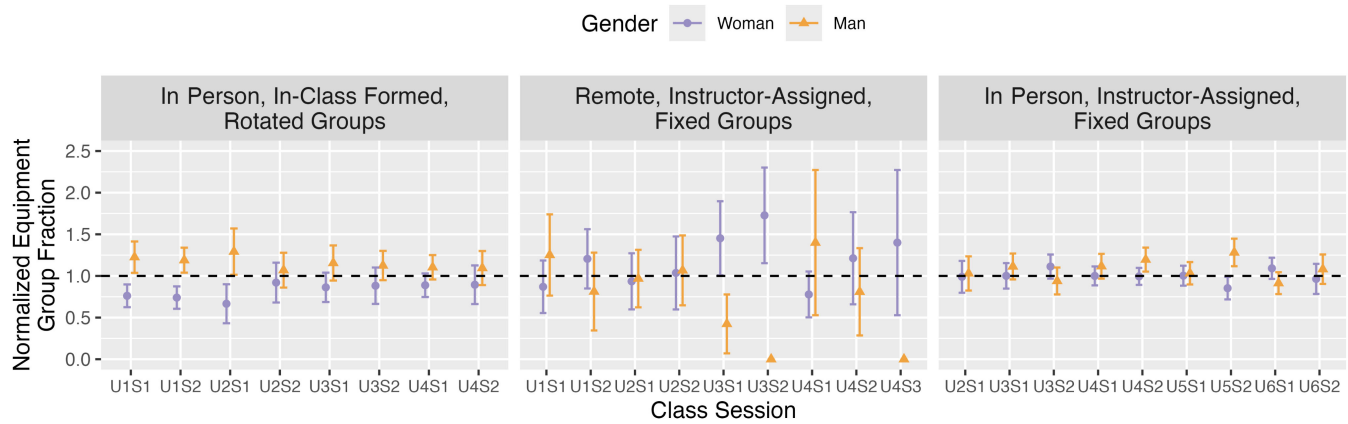


FIG. 2. The average of equipment group fraction for men and women normalized by group size across the semester in each course type. We normalized a student's group fraction from a class session by multiplying it by the number of students in their group. The dashed horizontal line indicates equitable equipment usage—a student in group size  $n$  is responsible for  $1/n$  of the group's equipment usage. The error bars represent the standard uncertainty of the mean. Each class session is labeled with its unit (U) and session (S).

do not fully determine which change in course structure (fixed vs rotated or instructor-assigned vs in-class formed) mitigated or supported inequities. It is possible that between instructor-assigned groups and groups held constant for the semester, one plays a lesser (or no) role in equipment equity—or that both changes play a role.

In our supplementary analysis of equipment individual fraction, we found generally the same trends as the main analysis with equipment group fraction. One difference, however, was that men spent slightly more of their class time than women using equipment in the in-person course with instructor-assigned, fixed groups—an inequity that was not observed in the equipment group fraction analysis. Our results suggest this discrepancy is likely due to small differences in equipment handling each class session that culminated in a small overall shift.

To better understand each group work structure variable's effect on group work equity, we investigated men and women's equipment group fraction across each class session of the semester. We accomplished this by comparing equipment group fractions (1) in each class session where students first join a new group and (2) across sessions where students are in the same groups.

By looking at the first session, we sought to isolate the effect of instructor-assigned vs in-class formed groups from the effect of fixed vs rotated groups. We assumed that the time effect associated with fixed vs rotated groups would not emerge until later in the course. In the case of in-class formed groups, we see that men used equipment more than women during the first session. For the in-person course with instructor-assigned groups, however, men and women have indistinguishable equipment group fractions in the first session. This suggests that instructor-assigned groups have a positive impact on equity, as gender differences were not present at the start of the semester before longer-term effects of rotated vs fixed groups could come into effect.

Next, we consider the role of fixed and rotated groups by looking at equipment group fraction over the semester. Men continue to have a greater average equipment group fraction every class session for the course with rotated groups, though the size of the difference decreases over time. The in-person course with fixed groups, however, has indistinguishable gender differences throughout most of the semester and, furthermore, women sometimes have a higher average equipment group fraction than men. The absence of consistently higher men's equipment group fraction suggests that fixed groups had a positive or neutral effect—neutral because we infer that the instructor-assigned groups created equitable participation in the first week, and so the fixed groups did not negatively impact that initial equity. The diminishing inequities in the rotated groups, however, suggests that rotated groups may in fact have a positive effect, in that the initial inequities became smaller over time. However, negative effects of rotated groups can also be seen at the start of each unit. That is, new groups were formed at the start of each unit, with most units lasting two weeks. The data seem to indicate that inequities are larger in the first week of each unit, just as new groups are formed, and diminish in the second week, after students have worked together already. Ultimately, this trend further speaks to the positive impacts of fixed groups.

Between these two trends, our results suggest that *both* instructors assigning groups and holding them fixed create a more equitable division of equipment usage within groups. This structure provides an alternative to that in Ref. [6], which found partner agreements remedied gender inequitable equipment usage in a course with in-class formed, rotated groups. Importantly, however, the effect of partner agreements was limited to the course with a lab practical exam, suggesting that students need a graded incentive to spend hands-on time with equipment. Only about one in eight lab courses, however, have lab

practical exams [23]. Our results, therefore, motivate using instructor-assigned groups that are fixed for the semester.

Our results also differed from one study that found no gender inequities in equipment usage with instructor-assigned—but rotated—groups [10]. The groups in their study, however, were pairs rather than groups of three to four. Additionally, they did not find gender inequitable equipment usage but instead found gender inequitable computer usage. It is possible that inequities from rotating groups manifested in our lab context with equipment usage and in their study's in computer usage. Future work should disaggregate equipment usage and other roles while studying the effects of group size and rotated vs fixed groups across course contexts.

These results provide some methods for instructors to consider when structuring lab group work so that students more equitably have hands-on time with equipment. Importantly, these alternative structures do not explicitly assign students roles, which students often disregard [32]. Direct manipulation of equipment, however, is not necessary for ICAP's higher levels of cognitive engagement: constructive and interactive [11]. It is possible that students, without directly handling equipment, still learn about equipment through these levels of engagement in other roles. For example, in a study of women in managerial and leadership roles in a physics lab, women used these roles to improve their understanding of the experiment as a whole [33]. In another study of women in physics labs, however, women in managerial and note-taker roles lost time on equipment and were prevented from gaining equipment expertise [9]. Future work should investigate the relationship between hands-on equipment use and cognitive engagement.

While we have explored the role of three variables, many more variables may be responsible for the effects observed here. We compared data with variables changed across several semesters, rather than within one semester with students randomly assigned to different instructional contexts. Thus, unobserved systematic differences in the instructional context or student population each semester may have covaried with our measured instructional variables, limiting our ability to claim causality in any case. Our observed students and groups are also not necessarily representative samples of each semester's course; too many students preferred not to share their gender for us to make this determination. Additionally, the remote semester's data was during the COVID-19 pandemic which could limit generalizability. While the course was designed so that any student could perform the experiments with materials found around their house or dorm, the *ease* of accessing such equipment might be different among students in dorm rooms vs those in their parents' homes, those in the USA vs those abroad, etc. In the future, universities providing remote lab courses will be more prepared than those forced to accommodate last minute changes due to the pandemic.

The results from this aspect of the study could help universities design group work for their remote courses. Alongside COVID-19, our studied courses are both before and after 2020, when students and universities gained increased awareness of diversity, equity, and inclusion efforts. This may have made students more inclined to include their classmates in equipment usage. The data were also taken at a single institution and so there are myriad geographical, institutional, and course context-specific factors that may or may not affect the generalizability of these results. For example, our course structure had groups of three to four students, which can create entirely different dynamics from partnerships as in other studies [6,9,10]—notably, groups of three or more can be made so that women students are not isolated. This research is also limited in context to a university with highly-motivated students who may be more likely to self-advocate for equipment usage; it is possible that these variables have different effects in different university contexts. Future work, therefore, should prioritize replication studies in other instructional contexts and further isolation of the variables tested and claims presented here.

## V. CONCLUSION

Comparing men's and women's equipment usage within their groups across three course contexts, we found that courses with instructor-assigned groups fixed for the semester exhibited reduced inequities in equipment usage. In our supplementary analysis, we find similar results; albeit, the in-person course with instructor-assigned groups fixed for the semester still featured small gender inequities. Given the number of studies finding a gender gap in equipment usage with few tested solutions, this change to group work provides an additional means to improving equity within introductory physics labs. To further understand the effectiveness of these two changes, future studies should investigate instructor-assigned groups and semester-long groups separately.

## ACKNOWLEDGMENTS

We would like to thank Ariel Baksh, Lauren Douglass, Samuel Karunwi, Lee Simpfendorfer, and Annie Tan for their help coding video data. We would also like to thank Anna McLean Phillips for helping ideate this work, Meagan Sundstrom and Scott Allen for helping record video data, and Jacob Feinleib for providing feedback in discussions of this work. We thank Matt Thomas and the Cornell Statistical Consulting Unit for their help with data analysis. Finally, we would like to thank Emma Hunt, Andrew Loveridge, Viranga Perera, Jonathan Perry, and Gregorio Ponti for discussions of this work. This material is based upon work supported by the National Science Foundation Grants No. DUE-1836617 and No. DGE-2139899, and the Cornell Center for Social Sciences.

**DATA AVAILABILITY**

The data that support the findings of this article are openly available [37].

**APPENDIX A: COURSE DEMOGRAPHICS**

The demographics for each semester's offering of the lab course can be found in Table V. Notably, these courses are approximately 50/50 men and women.

TABLE V. Course-level student demographics from each semester. Responses were from a pre- and post-semester survey. The in person, in-class formed, rotated groups survey had a 98.6% completion rate; the remote, instructor-assigned, fixed groups survey had a 95.3% completion rate; and the in person, instructor-assigned, fixed groups survey had a 91.7% completion rate.

| Student-level variables                   | Survey responses                           |                                           |                                              |
|-------------------------------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
|                                           | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
| All                                       | 219                                        | 407                                       | 600                                          |
| Gender                                    |                                            |                                           |                                              |
| Woman                                     | 119                                        | 172                                       | 293                                          |
| Man                                       | 95                                         | 156                                       | 256                                          |
| Nonbinary or other                        | 3                                          | 5                                         | 6                                            |
| Did not disclose                          | 2                                          | 74                                        | 45                                           |
| Race or ethnicity                         |                                            |                                           |                                              |
| American Indian or Alaska Native          | 2                                          | 4                                         | 6                                            |
| Asian                                     | 95                                         | 109                                       | 206                                          |
| Black or African American                 | 14                                         | 36                                        | 58                                           |
| Hispanic or Latino                        | 26                                         | 56                                        | 73                                           |
| Native Hawaiian or Other Pacific Islander | 0                                          | 2                                         | 2                                            |
| White                                     | 101                                        | 151                                       | 245                                          |
| Some other race or ethnicity              | 4                                          | 5                                         | 10                                           |
| Did not disclose                          | 12                                         | 131                                       | 97                                           |
| Class standing                            |                                            |                                           |                                              |
| Freshman                                  | 163                                        | 316                                       | 549                                          |
| Sophomore                                 | 47                                         | 13                                        | 15                                           |
| Junior                                    | 6                                          | 2                                         | 2                                            |
| Senior                                    | 1                                          | 2                                         | 2                                            |
| Did not disclose                          | 2                                          | 74                                        | 32                                           |

(Table continued)

TABLE V. (Continued)

| Student-level variables                    | Survey responses                           |                                           |                                              |
|--------------------------------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
|                                            | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
| Major                                      |                                            |                                           |                                              |
| Physics, astronomy, or engineering physics | 9                                          | 15                                        | 45                                           |
| Engineering                                | 173                                        | 269                                       | 442                                          |
| Life science or biology                    | 6                                          | 13                                        | 5                                            |
| Other physical science                     | 22                                         | 16                                        | 32                                           |
| Other                                      | 6                                          | 5                                         | 8                                            |
| Did not disclose                           | 3                                          | 89                                        | 68                                           |

**APPENDIX B: ASSUMPTIONS FOR HIERARCHICAL LINEAR MODELING**

Here, we discuss our model's ability to meet hierarchical linear modeling's assumptions. We check assumptions according to advice from Van Dusen and Nissen [31]. We check for the assumption of linearity in Fig. 3. Although we see diagonal lines at the bottom left and top right of each plot, these are due to upper and lower bounds on our data (a student cannot have an equipment group fraction less than 0 or greater than 1). Otherwise, there are no discernible trends. For the assumption of homogeneity of variance, our ANOVA returned no significant results (the in person, in-class formed, rotated groups course returned  $p \sim 1$ ; the remote, instructor-assigned, fixed groups course returned  $p = 0.8577$ ; and the in person, instructor-assigned, fixed groups course returned  $p \sim 1$ ), indicating homogeneity of variance. Finally, for the assumption of normality, we plotted our

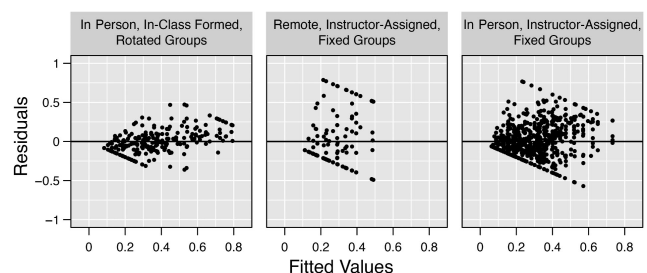


FIG. 3. A visual check for hierarchical linear modeling's assumption of linearity for the group fraction model. These plots display the residuals vs fitted values for each course.

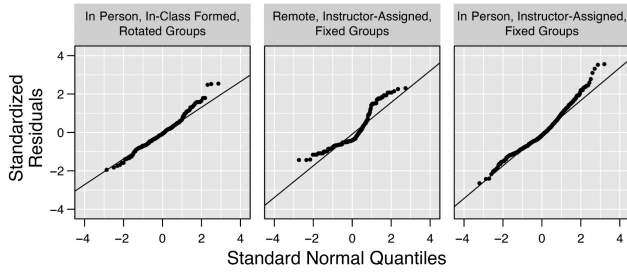


FIG. 4. A visual check for hierarchical linear modeling’s assumption of normality for the group fraction model. These plots display the residuals vs fitted values for each course.

residuals in Fig. 4. For the in person, in-class formed, rotated groups course and the in person, instructor-assigned, fixed groups course, our data are on or near the line. Though the data stray from the line at the far ends, this is generally acceptable [34,35]. The data for the remote, instructor-assigned, fixed groups course sticks to the line less consistently, suggesting the regression results should be interpreted with caution. These data were included primarily to provide additional context and a comparison for the in-person course with instructor-assigned, fixed groups and so the regression fits do not significantly contribute to the major claims.

We used the `lmer` function from the `LME4` package in R to run our hierarchical linear models [36].

## APPENDIX C: EQUIPMENT INDIVIDUAL FRACTION

Here, we discuss an alternate methodology for assessing equipment usage: equipment individual fraction. This quantity evaluates how individual students are using time each class session regardless of the behavior of their group mates. Within this section, we will focus only on how these results differ from those presented using equipment group fraction in the main text. In Appendix C 1, we discuss our methodology. In Appendix C 2, we present our results and discuss them in Appendix C 3.

### 1. Methods

Equipment group fraction allows for effective normalization across labs that involve different amounts of equipment usage. It may dilute effects, however, by treating groups (and therefore individual students) that spent substantially more time on equipment equivalent to groups that barely used equipment. We use here an alternate measure that considers the amount of time each individual student spends on equipment.

Because students may leave lab at different times, we normalized equipment observations by the total observed instances in the lab. We refer to this as “equipment individual fraction.” The equipment individual fraction

for student  $i$  in the  $j$ th class session,  $f_{ij}$ , is given in Eq. (C1),

$$f_{ij} = \frac{N_{ij}}{M_{ij}}, \quad (\text{C1})$$

where  $N_{ij}$  is the number of observations of student  $i$  using equipment in class session  $j$ , and  $M_{ij}$  is the total number of observations of student  $i$  in the  $j$ th class session.

Here, we replicate the analysis in the main text (which was on equipment group fraction) with individual fraction. One change to the regression models is that we do not include the `GroupSize` term in the level-1 data. For group fraction, more group members means an equitable distribution is inherently smaller. This is not the case for individual fraction. Therefore, we do not include this term in our model.

To again better understand the differences between fixed/rotated and instructor-assigned/in-class formed groups, we investigated men and women’s equipment individual fraction across class sessions. As above, we do not normalize our data by group size and only plot the average equipment individual fraction.

We additionally perform another regression model that considers group gender composition. Comparing average equipment group fraction for men in all-men groups to women in all-women groups is uninformative as the average usage by gender is the same. For a group of  $n$  students, the average equipment group fraction is always  $1/n$ . For mixed-gender groups, this means the average equipment group fraction for men and women varies, but for single-gender groups, the average equipment group fraction for men and women is always  $1/n$ .

Consider, for example, a group of three students: two women and one man, where the first woman is observed handling equipment 6 times, the second woman is observed handling the equipment 4 times, and the man is not observed handling the equipment. Their equipment group fractions would be  $6/10$ ,  $4/10$ , and  $0/10$ , respectively, and the average of the group’s equipment group fractions would be  $1/3$  ( $1/n$ ). The average of the women’s equipment group fractions would be  $1/2$  and the average of the men’s equipment group fractions would be  $0$ . If that same group were all women, however, the average of the women’s equipment group fractions would necessarily be  $1/3$  ( $1/n$ ). Thus, equipment group fraction can primarily distinguish the participation of students in mixed-gender groups.

Equipment individual fraction, however, allows us to compare men and women’s class time spent on equipment in both single-gender and mixed-gender groups. We add one term to the original regression’s level-2 data as in Eq. (C2),

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Gender} + \gamma_{02}\text{GenderComp} + u_{0j}, \quad (\text{C2})$$

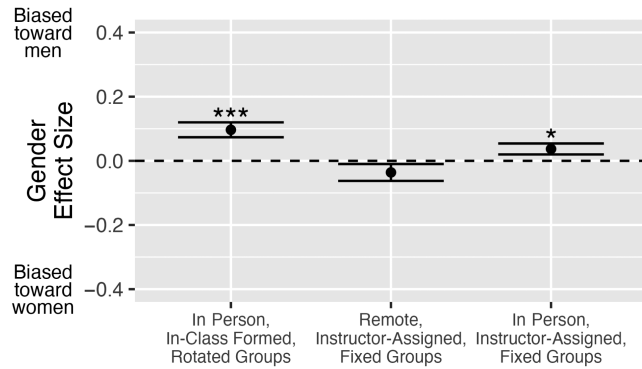


FIG. 5. Results from multilevel regression for equipment individual fraction controlling for random effects. A value greater than zero indicates men spent more of their class time on equipment than women, whereas a value less than zero indicates less. The error bars represent the standard error of the regression coefficients. The asterisks denote statistical significance where \* indicates  $p < 0.05$  and \*\*\* indicates  $p < 0.001$ . The full model output can be found in Table VI.

where GenderComp is a binary term indicating whether the student was in a group that was single gender or mixed gender. The interaction term with gender allows us to distinguish all-men groups from all-women groups. Groups with nonbinary students that otherwise had a single-gender composition were coded as mixed gender. Because we are interested in comparing how men and women use equipment in each group type, we use *post hoc* contrasts (denoted by  $\Delta$ ) to compare to our base term: women in mixed-gender groups.

We include checks for our assumptions for hierarchical linear modeling in Supplemental Material [37]. These checks resemble the results for the main text analysis outlined in Appendix B.

## 2. Results

The results from the equipment individual fraction regression are shown in Fig. 5 and Table VI. The primary difference between this analysis and the main text equipment group fraction analysis is that—in addition to the in-person course with in-class formed, rotated groups—men also spent more of their class time on equipment than

women for the in-person course with instructor-assigned, fixed groups ( $\gamma_{01} = 0.037 \pm 0.017$ ,  $p = 0.030$ ), controlling for random effects.

We include the results separated by class session for each course in Fig. 6. For the in-person course with in-class formed, rotated groups, compared to equipment group fraction, men again have a higher measure of equipment usage across the semester, but the difference between men and women does not decrease over the semester. Additionally, the pattern of decreased differences between men and women in the first and second sessions is similar to that seen with equipment group fraction. For the remote course with instructor-assigned, fixed groups, the results are similar to those with equipment group fraction. For the in-person course with instructor-assigned, fixed groups, we see no differences at the start of the semester much like equipment group fraction. There are more class sessions, however, where men have a greater measure of equipment usage than woman, as compared to the main text analysis.

We also analyze the individual fraction data based on group gender composition (Table VII and Fig. 7). For the in-person course with in-class formed, rotated groups, women in single-gender groups did not spend more of their class time on equipment than women in mixed-gender groups ( $\Delta = -0.018 \pm 0.033$ ,  $p = 0.596$ ). Men in both single-gender groups ( $\Delta = 0.091 \pm 0.034$ ,  $p = 0.009$ ) and mixed-gender groups ( $\Delta = 0.088 \pm 0.031$ ,  $p = 0.005$ ) spent more time on equipment than women in mixed-gender groups. In both of the courses with instructor-assigned, fixed groups, we see no statistically significant differences in time spent on equipment between men and women controlling for group gender composition.

## 3. Discussion

When we instead look at students' individual equipment usage each lab, rather than within their groups, we replicate many of the main text findings through equipment group fraction. Again, we find that men spent more of their class time using equipment than women in the in-person course with in-class formed, rotated groups. We also find no statistically significant difference between men and women in class time spent on equipment for the remote course with instructor-assigned formed groups fixed all semester.

TABLE VI. Results from linear regression for equipment individual fraction. The table includes the regression coefficients, standard errors, and  $p$  values in parentheses. The conditional (marginal)  $R^2$  values for these models are 0.567 (0.091) for in person, in-class formed, rotated groups; 0.080 (0.015) for remote, instructor-assigned, fixed groups; and 0.353 (0.014) for in person, instructor-assigned, fixed groups.

| Predictor           | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
|---------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
| Intercept           | $0.139 \pm 0.016$ ( $<0.001$ )             | $0.116 \pm 0.017$ ( $<0.001$ )            | $0.147 \pm 0.011$ ( $<0.001$ )               |
| Gender (ref: women) |                                            |                                           |                                              |
| Men                 | $0.097 \pm 0.023$ ( $<0.001$ )             | $-0.036 \pm 0.026$ (0.171)                | $0.037 \pm 0.017$ (0.030)                    |

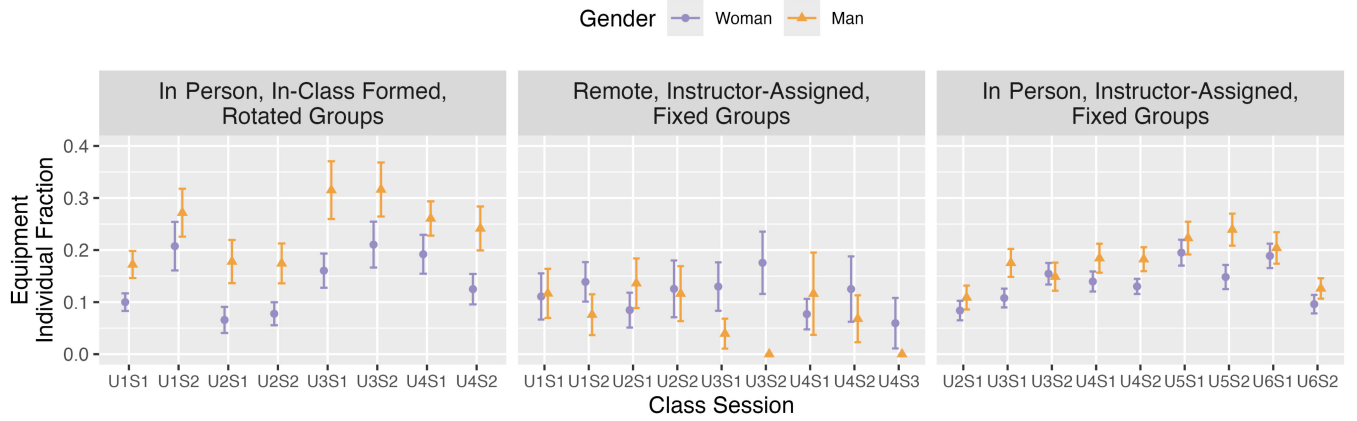


FIG. 6. The average of equipment individual fraction for men and women across the semester in each course type. The error bars represent the standard uncertainty of the mean.

Contrary to our group fraction results, however, we find that men spent more class time on equipment than women in the in-person course with instructor-assigned, fixed groups. We consider two possible reasons for this difference in results: men in single-gender groups spent more time on equipment than men in mixed-gender groups and/or groups spent more time on equipment when men were their group's equipment user.

For the first explanation, we restate—as detailed in the Appendix C 1—that equipment group fraction cannot distinguish the average participation of men and women in single-gender groups. If men in single-gender groups spent more time on equipment than other students, this would produce an overall larger equipment individual fraction but not a larger equipment group fraction for men compared with women. For both in-person courses, however, men in single- and mixed-gender groups have similar average equipment individual fractions. This contrasts another study of an in-person course with in-class formed, rotated groups that found men in single-gender groups spent more time on equipment than men in mixed-gender groups [7]. We also find that men in both single- and mixed-gender groups have a slightly higher average

individual fraction than women in mixed-gender groups (this difference is smaller in the in-person course with intentional fixed groups than in the in person, in-class formed, rotated groups and not statistically significant in the former). This suggests that the differences we see between men's and women's equipment individual fraction in Fig. 5 is due to a combination of smaller effects across all group gender compositions.

The second possible explanation for our difference between equipment group fraction and equipment individual fraction results is that when men were their group's equipment user, groups spent more class time on equipment than in other sessions. In our data, this can be discerned by comparing Figs. 2 and 6. Were this explanation true, we would see that students' equipment individual fraction is particularly high (i.e., above that of other class sessions) for class sessions where men's equipment group fraction is greater than women's (orange dots above the purple dots). For the in-person course with instructor-assigned, fixed groups, we instead find that men have a greater average equipment group fraction than women in every class session except for U3S2, U5S2, and U6S1. There does not appear to be an obvious trend between the size of the

TABLE VII. Results from linear regression for equipment individual fraction with group gender composition. The table includes the regression coefficients, standard errors, and  $p$  values in parentheses. The conditional (marginal)  $R^2$  values for these models are 0.571 (0.092) for in person, in-class formed, rotated groups; 0.086 (0.015) for remote, instructor-assigned, fixed groups; and 0.335 (0.013) for in person, instructor-assigned, fixed groups.

| Predictor                                    | In person, in-class formed, rotated groups | Remote, instructor-assigned, fixed groups | In person, instructor-assigned, fixed groups |
|----------------------------------------------|--------------------------------------------|-------------------------------------------|----------------------------------------------|
| Intercept                                    | $0.147 \pm 0.022$ ( $<0.001$ )             | $0.120 \pm 0.020$ ( $<0.001$ )            | $0.151 \pm 0.014$ ( $<0.001$ )               |
| Gender (ref: women)                          |                                            |                                           |                                              |
| Men                                          | $0.088 \pm 0.031$ (0.005)                  | $-0.040 \pm 0.029$ (0.166)                | $0.028 \pm 0.022$ (0.196)                    |
| Group gender composition (ref: mixed gender) |                                            |                                           |                                              |
| Women, single gender                         | $-0.018 \pm 0.033$ (0.596)                 | $-0.014 \pm 0.037$ (0.705)                | $-0.013 \pm 0.022$ (0.558)                   |
| Men, single gender                           | $0.021 \pm 0.047$ (0.660)                  |                                           | $0.015 \pm 0.035$ (0.664)                    |

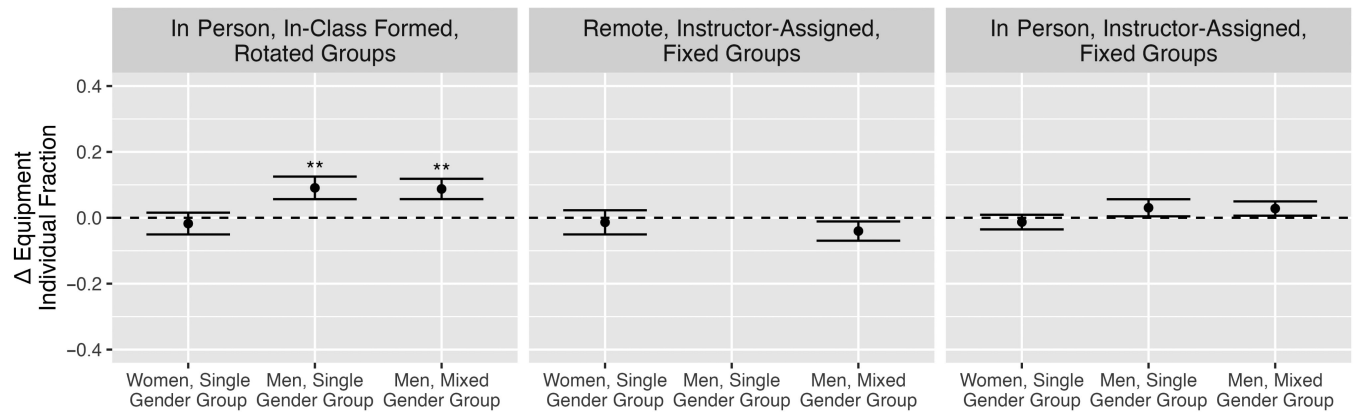


FIG. 7. Results from *post hoc* contrasts following the hierarchical linear model for equipment individual fraction described in Eq. (C2). The base term is women in mixed-gender groups. The error bars represent the standard error of the difference in means. The asterisks denote statistical significance where \*\* indicates  $p < 0.01$ . The full linear regression model output can be found in Table VII. We do not plot any data for men in single-gender groups for the remote course with instructor-assigned, fixed groups because there were no all-men groups in that course.

difference in men and women's fractions and the overall average equipment usage.

Finally, the analysis of individual group fraction across lab sessions replicated the analysis in the main text. The only exception was that the difference between men's and women's participation did not decrease over time within the in-person course with in-class formed, rotated groups. This

result lends additional evidence to the claim in the main text that keeping groups fixed mitigated differences in men's and women's participation. While differences decreasing over time for equipment group fraction suggested the course was potentially becoming more equitable with time, the static differences for equipment individual fraction suggest otherwise.

- [1] J. Kozminski, N. Beverly, D. Deardorff, R. Dietz, M. Eblen-Zayas, R. Hobbs, H. Lewandowski, S. Lindaas, A. Reagan, R. Tagg, J. Williams, and B. Zwickl, AAPT recommendations for the undergraduate physics laboratory curriculum, AAPT Technical Report, 2014, [https://www.aapt.org/resources/upload/labguidelinesdocument\\_ebendorsed\\_nov10.pdf](https://www.aapt.org/resources/upload/labguidelinesdocument_ebendorsed_nov10.pdf).
- [2] N. G. Holmes, G. Heath, K. Hubenig, S. Jeon, Z. Y. Kalender, E. Stump, and E. C. Sayre, Evaluating the role of student preference in physics lab group equity, *Phys. Rev. Phys. Educ. Res.* **18**, 010106 (2022).
- [3] D. Doucette and C. Singh, Making lab group work equitable and inclusive, *J. Coll. Sci. Teach.* **52**, 31 (2023).
- [4] P. W. Laws, P. J. Rosborough, and F. J. Poodry, Women's responses to an activity-based introductory physics program, *Am. J. Phys.* **67**, S32 (1999).
- [5] M. Dew, A. Phillips, S. Karunwi, A. Baksh, E. M. Stump, and N. G. Holmes, So unfair it's fair: Equipment handling in remote versus in-person introductory physics labs, presented at PER Conf. 2022, Grand Rapids, MI, 10.1119/perc.2022.pr.Dew, pp. 144–150.
- [6] M. Dew, E. Hunt, V. Perera, J. Perry, G. Ponti, and A. Loveridge, Group dynamics in inquiry-based labs: Gender inequities and the efficacy of partner agreements, *Phys. Rev. Phys. Educ. Res.* **20**, 010121 (2024).
- [7] K. N. Quinn, M. M. Kelley, K. L. McGill, E. M. Smith, Z. Whipps, and N. G. Holmes, Group roles in unstructured labs show inequitable gender divide, *Phys. Rev. Phys. Educ. Res.* **16**, 010129 (2020).
- [8] R. Gunter, G. Spiczak, and J. Madsen, Cosmic collaboration in an undergraduate astrophysics laboratory, *Am. J. Phys.* **78**, 1035 (2010).
- [9] D. Doucette, R. Clark, and C. Singh, Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning, *Eur. J. Phys.* **41**, 035702 (2020).
- [10] J. Day, J. B. Stang, N. G. Holmes, D. Kumar, and D. A. Bonn, Gender gaps and gendered action in a first-year physics laboratory, *Phys. Rev. Phys. Educ. Res.* **12**, 020104 (2016).
- [11] M. T. H. Chi and R. Wylie, The ICAP framework: Linking cognitive engagement to active learning outcomes, *Educ. Psychol.* **49**, 219 (2014).
- [12] J. Jovanovic and S. S. King, Boys and girls in the performance-based science classroom: Who's doing the performing?, *Am. Educ. Res. J.* **35**, 477 (1998).
- [13] T. A. Greenfield, Gender and grade-level differences in science interest and participation, *Sci. Educ.* **81**, 259 (1997).

- [14] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [15] K.D. Tanner, Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity, *Cell Biol. Educ.* **12**, 322 (2013).
- [16] D. Doucette and C. Singh, Share it, don't split it: Can equitable group work improve student outcomes?, *Phys. Teach.* **60**, 166 (2022).
- [17] D. Doucette, R. Clark, and C. Singh, What makes a good physics lab partner?, presented at PER Conf. Proc. 2020, [10.1119/perc.2020.pr.Doucette](https://doi.org/10.1119/perc.2020.pr.Doucette), pp. 124–130.
- [18] J. P. Adams, G. Brissenden, R. S. Lindell, T. F. Slater, and J. Wallace, Observations of student behavior in collaborative learning groups, *Astron. Educ. Rev.* **1**, 25 (2002).
- [19] D. W. Johnson and R. T. Johnson, An educational psychology success story: Social interdependence theory and cooperative learning, *Educ. Res.* **38**, 365 (2009).
- [20] D. W. Johnson, R. T. Johnson, and K. A. Smith, Cooperative learning returns to college: What evidence is there that it works?, *Change* **30**, 26 (1998).
- [21] K. Tanner and D. Allen, Approaches to biology teaching and learning: Understanding the wrong answers—teaching toward conceptual change, *Cell Biol. Educ.* **4**, 112 (2005).
- [22] D. W. Johnson and R. T. Johnson, Making cooperative learning work, *Theory Pract.* **38**, 67 (1999).
- [23] G. Geschwind, M. Alemani, M. F. J. Fox, P. S. W. M. Logman, E. Tufino, and H. J. Lewandowski, Development of a global landscape of undergraduate physics laboratory courses, *Phys. Rev. Phys. Educ. Res.* **20**, 020117 (2024).
- [24] N. Shah and C. M. Lewis, Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework, *Cognit. Instr.* **37**, 423 (2019).
- [25] N. G. Holmes, Structured quantitative inquiry labs: Developing critical thinking in the introductory physics laboratory, Ph.D. thesis, University of British Columbia, 2014.
- [26] D. L. Schwartz and T. Martin, Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction, *Cognit. Instr.* **22**, 129 (2004).
- [27] R. K. Atkinson, S. J. Derry, A. Renkl, and D. Wortham, Learning from examples: Instructional principles from the worked examples research, *Rev. Educ. Res.* **70**, 181 (2000).
- [28] J. L. Fleiss, Measuring nominal scale agreement among many raters, *Psychol. Bull.* **76**, 378 (1971).
- [29] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed. (Wiley, New York, 2003).
- [30] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, *Phys. Rev. Phys. Educ. Res.* **12**, 020114 (2016).
- [31] B. Van Dusen and J. Nissen, Modernizing use of regression models in physics education research: A review of hierarchical linear modeling, *Phys. Rev. Phys. Educ. Res.* **15**, 020108 (2019).
- [32] Y. Chang and P. Brickman, When group work doesn't work: Insights from students, *CBE-Life Sci. Educ.* **17**, ar52 (2018).
- [33] E. M. Stump and N. G. Holmes, Student views of what counts as doing physics in the lab, presented at PER Proc. Conf. 2022, Grand Rapids, MI, [10.1119/perc.2022.pr.-Stump](https://doi.org/10.1119/perc.2022.pr.-Stump), pp. 444–450.
- [34] T. Lumley, P. Diehr, S. Emerson, and L. Chen, The importance of the normality assumption in large public health data sets, *Annu. Rev. Public Health* **23**, 151 (2002).
- [35] C. Walsh, H. J. Lewandowski, and N. G. Holmes, Skills-focused lab instruction improves critical thinking skills and experimentation views for all students, *Phys. Rev. Phys. Educ. Res.* **18**, 010128 (2022).
- [36] D. Bates, M. Mächler, B. Bolker, and S. Walker, Fitting linear mixed-effects models using LME4, *J. Stat. Softw.* **67**, 1 (2015).
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/yyml-sr6q> for the observation data, code used for analysis, and code output.