

# Characterizing studio physics instruction across instructors and institutions

Matthew Wilcox<sup>1</sup>, Gerald Feldman<sup>2</sup>, Joshua Von Korff<sup>3</sup>, Noel Klingler<sup>2</sup>, Ozden Sengul<sup>3</sup>, and Jacquelyn J. Chini<sup>1</sup>

<sup>1</sup>University of Central Florida, Department of Physics, 4111 Libra Drive, Orlando, FL, Orlando, FL, 32816

<sup>2</sup>George Washington University, Department of Physics, 725 21<sup>st</sup> Street NW, Washington, DC, 20052

<sup>3</sup>Georgia State University, Department of Physics and Astronomy, 25 Park Place, Atlanta, GA, 30303

An increasing number of institutions are adopting a collaborative student-centered studio approach for their introductory physics classes, although there is considerable variation in their deployments and a wide range of success in these different cases. Using a modified version of the TDOP observational protocol, we observed and coded 13 instructors teaching SCALE-UP (one studio implementation method) physics classes at two universities to characterize each studio class. We coded different types of instructor dialogue, class discussion, and students' group and individual work, as well as technology used in the classroom. We identified both similarities and differences among the various classes. Here, we report the percentage of intervals in which certain codes were observed, highlighting the most prevalent codes and noting common code combinations. This is the beginning of work to characterize different studio physics classes to determine effective practices.

## I. INTRODUCTION

We define introductory physics studio courses as those that combine components, such as lecture, laboratory and recitation, which are traditionally separated in a traditional course. Studio courses typically emphasize student group work, and reduced faculty lecture. One popular mode of studio for introductory physics is SCALE-UP, created by Beichner and colleagues at North Carolina State University (NCSSU) [1]. This SCALE-UP model has been shown to improve learning gains and reduce the failure rates of underrepresented students [1]. SCALE-UP was designed to integrate lecture, laboratory and recitation components into a single classroom environment with 4-6 hours of activity-based instruction per week for large enrollment classes [2]. Students sit at round tables with access to computers and laboratory equipment that support them to learn physics through activities done by small groups [1]. In class, technology may be used to collect and analyze experimental data, run simulations, or construct mathematical models of physical situations [1,3].

Among active-learning strategies in physics, SCALE-UP is unusual in its flexibility. The SCALE-UP pedagogy encourages three broad goals (cooperative learning, minimized lecture, and focus on student reasoning and presentation) rather than prescribing a particular curriculum [4]. There is typically some amount of lecture, even if only a few minutes at a time, but the extent of the lecture can vary. SCALE-UP need not include clickers, although it often does. There is generally a laboratory component to SCALE-UP, but the degree of emphasis on the labs can vary, and there are no set texts or activities. The labs may range from short "tangibles" to more traditional two-hour sessions with lab reports. The amount of lecture, lab, and other activities, as well as factors like instructor experience may vary largely between classrooms. Thus, it may not be surprising that the learning gains associated with SCALE-UP, such as conceptual understanding and attitudes, also vary [2, 5]. One possible explanation for the variation in learning gains comes

from differences in teaching practices at the levels of institution and instructor.

How, then, can we know what kinds of studio teaching practices may be responsible for student learning? Foote and colleagues have catalogued the use of SCALE-UP by several different instructors and institutions [6] with a more recent focus on curriculum development [4]. The fundamental problem is that there has been little empirical work to characterize how studio and SCALE-UP classes are actually taught in classrooms across the country. One of the driving goals of this project is to address this issue. In this paper, we present our initial findings on the characterization of SCALE-UP-type classes at two large, research-intensive universities. While the overall goal of our research project is to investigate studio classes, in this paper we focus on the SCALE-UP model as this is the model that has the most similarity with the classes observed.

## II. SETTING AND PARTICIPANTS

At University A, there are currently two sections each of four studio courses per semester, each taught by a separate instructor. Instructors have the freedom to teach studio in

**TABLE 1.** A comparison of Universities A and B.

University A	University B
99 student capacity	81 student capacity
Both SCALE-UP and traditional courses offered.	All introductory courses are SCALE-UP.
2 full professors 3 lecturers 2 assistant professors 1 associate professor	2 full professors 4 adjunct faculty
SCALE-UP teaching experience: <ul style="list-style-type: none"><li>• As low as 2 semesters</li><li>• Up to 10+ semesters</li><li>• Most with 5+ semesters experience.</li></ul>	SCALE-UP teaching experience: <ul style="list-style-type: none"><li>• As low as no prior experience</li><li>• Up to 7+ semesters</li><li>• Most with 4+ semesters</li></ul>

their own way and have experimented with several different curricula; internal assessment suggests learning gains vary widely across individual sections. At University B, there are eight sections of physics each semester. At the time of the site visit, six instructors (two full professors and four adjunct faculty) were teaching six sections of introductory physics. Table 1 compares the settings and participants for Universities A and B. Of note is that both traditional and SCALE-UP classes are offered at University A, while all introductory physics courses were converted to SCALE-UP several years ago at University B.

### III. METHODS

#### A. Coding scheme

To record and measure actions occurring in studio classrooms, we created and used a modified version of the Teaching Dimensions Observation Protocol (TDOP) [7]. TDOP records actions in two-minute intervals. Our changes to TDOP included the removal of codes not applicable to studio classes or our broader objectives, the modification and elaboration of code definitions to enhance clarification for our setting and team, and the creation of new codes to describe typical studio activities.

We reorganized the codes into categories that reflect characteristic phases of the typical studio-style course. The "One Conversation" category is applicable when only one person is speaking in the class at a time, and captures instructor actions such as lecturing, giving feedback to students, posing a content-based question, or conducting a demonstration, and also records student actions such as asking the instructor a question, responding to an instructor question, engaging in a class-wide discussion, or delivering a presentation. The "Parallel Work" category captures class-wide activity such as small group work, individual work, instructor/TA interactions with a group or individual, and assessments. The "Parallel Response" category gauges Peer Instruction-type periods, such as multiple-choice questions, conceptual prediction questions, or numerical calculations. Lastly, the "Technology" category records the objects used in the above categories (e.g., lab equipment, books, static displays/PowerPoint slides, clickers, worksheets, etc.), and also where they occurred (e.g., projector screen, classroom wall/whiteboard, student tables, etc.). As studio interactions tend to periodically shift between distinct periods of one conversation and active engagement, this categorization of codes was designed not only because of the similarity of the codes within them, but also to enhance inter-rater reliability (IRR) by grouping together commonly coincident codes.

We will specifically discuss four codes that represent many of the actions and tools important in describing how SCALE-UP differs from traditional physics classes. These codes are for lecture, feedback, clicker questions, and small group work.

When an instructor addressed a class, we coded it as either lecture or feedback. Feedback is similar to lecture, though it requires the instructor to refer to something students have previously worked on. Clickers were coded when an instructor gave the students a question to answer using clicker technology. Small group work was coded when students formed groups of two or more for the purpose of discussion or completion of a task.

#### B. Inter-rater reliability

Before collecting data, we needed to establish sufficient IRR between our observers. The first step in that process was to watch 20-30 minute videos of a SCALE-UP classroom while coding individually. Afterwards, observers met to discuss and resolve disagreements. Then, two multi-day trips were made to Georgia State University for coders to practice coding in real-time situations, to explore how well the codes fit, and to discuss code meanings; our external evaluator was present on the first trip to make sure our codes would make sense to those outside the project. Once all observers consistently achieved satisfactory IRR on the practice videos and felt comfortable with the coding scheme, they met in-person at the University of Central Florida to practice live coding. The four observers all individually coded four different instructors teaching a SCALE-UP class for one hour at a time. Immediately following the coding sessions, they again discussed any disagreements to resolve any problems.

IRR was measured on a code-by-code basis and then a single value corresponding to all observers' IRR was calculated. We used Krippendorff's Alpha as our measure for IRR because we wanted to continuously measure IRR throughout the data collection, and Krippendorff's Alpha can handle missing cases (for example, when one observer is not present for a class). An alpha value greater than or equal to 0.8 means the observers have achieved a high level of confidence in IRR for that code, while a value of 0.67 or higher indicates agreement with which one may draw only tentative conclusions [8]. To calculate the single value Krippendorff's Alpha, we found the weighted average of all the codes' alpha values, weighting by the number of intervals in which each code was observed. During the training phase, the four observers achieved a weighted average Krippendorff's Alpha of 0.67. During the data collection phase, which was all live coding, the observers achieved an average IRR of 0.78. Throughout the observations the observers achieved satisfactory IRR on many, but not all, of the individual codes.

**TABLE 2.** Sample TDOP codes and the associated Krippendorff's Alpha.

Code	Lecture	Feedback	Clicker	Small Group Work
Alpha	0.85	0.80	0.89	0.86

## IV. RESULTS AND ANALYSIS

### A. Inter-rater reliability

There were many codes in which satisfactory IRR was achieved. Table 2 shows the Krippendorff's Alpha values for the codes we are specifically discussing in this paper.

Two codes that may be important for a SCALE-UP class that we did not achieve sufficient IRR on were the demonstration and student presentation codes. The alpha values for these codes were 0.64 and 0.73, respectively. One possible explanation for the low level of IRR for these codes is that they occurred infrequently. Out of 652 total intervals observed, at least one of the observers coded demonstration in 55 intervals and student presentation in only 11. When a code occurs so infrequently, random chance plays a greater role in IRR. While the observers did not achieve a satisfactory IRR on these codes, it is important to note how infrequently they occurred for codes that one might expect to occur often in a SCALE-UP classroom.

### B. Code frequencies

#### 1. Individual codes

At University A, lecture occurred in 22% of all intervals. Similarly, at University B, lecture occurred in 20% of all intervals. Within each university, the distribution of lecture use is similar as well (see Fig. 1a). At both universities, the majority of instructors were observed to lecture in 10-30% of the intervals. One instructor at each university lectured at around 40% of the intervals and one instructor at each lectured in less than 10% of the intervals. At University A, lecture use ranges from 7% to 43%, and at University B lecture use ranges from 7% to 36%.

Feedback was observed more often than lecture at both universities. At University A, feedback was observed in 23% of the intervals, and at University B feedback was observed in 30%. The distributions within each university are fairly similar as well (see Fig. 1b). The distributions of Universities A and University B are nearly identical except for the two extra instructors at University A who either spent less than

10% of the intervals using feedback or between 40% and 50%. Both universities have a fairly broad distribution of feedback use ranging from 10% to 52% for University B and slightly broader at University A ranging from 0% to 52%. Overall, either lecture or feedback occurred in 49% of the intervals.

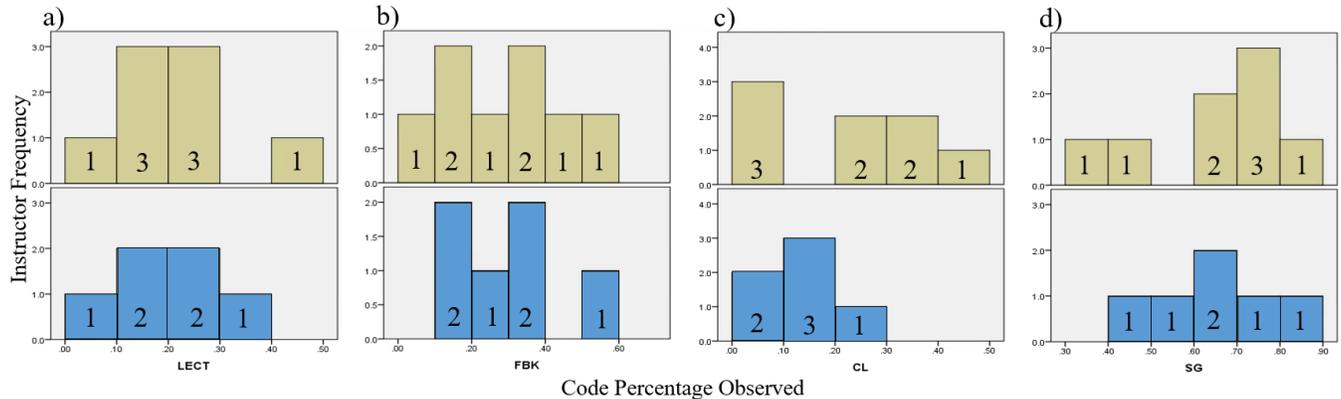
The use of clicker technology varied largely at University A. Three instructors used clickers in less than 10% of the observed intervals while the other five instructors used clickers in 20-44% of the intervals. At University B, clicker use was more consistent (see Fig. 1c). Instructors at University B used clickers in an average of 11% of the intervals with a range between 0% and 22%. Compare that to University A, where instructors used clickers in an average of 21% of the intervals with a range between 0% and 44%. Due to the large variation at University A, there is not a statistically significant difference in the means for clicker use between the two universities.

In general, small group work occurred in a majority of all intervals observed. At University A and University B, the average use of small group work was 66% and 63% respectively. However, the distributions at each university appear to be different (see Fig. 1d). At University A, many instructors spent 66-79% doing small group work. Two outliers at 36% and 40% bring the University A average down to 66%. At University B we find a much more even distribution of small group work, with instructors spending 42-83% of the intervals engaging in that activity.

#### 2. Coding combinations

As part of our analysis, we looked at frequently occurring codes with high IRR that commonly occurred together. We find two sets of common coding combinations.

First, across all observations, when clicker was coded, small group was also coded 93% of the time. At University A, small group was observed 92% of the time that clicker was coded and at University B this number is 94%. This shows that clickers are used almost exclusively in a small group work setting. This heavy use of small group clicker questions may go against the Peer Instruction model



**FIG 1.** Frequency of instructors (y-axis) for University A (gold) and University B (blue) in which Lecture (a), Feedback (b), Clickers (c), and Small Group Work (d) were observed in a percentage of intervals (x-axis).

involving a two-step process of first answering a clicker question individually, then discussing and answering again with a small group.

Secondly, we notice that both instructors and graduate/undergraduate teaching assistants are very interactive during small group work. Overall, when small group work was coded, a TA interaction was also coded in 72% of those intervals. Similarly, in 62% of the small group codes, an instructor interaction was also observed. At University A, these values are 76% for TAs and 56% for instructors. At University B, the values are 66% for TAs and 70% for instructors. We see this as a favorable observation because one important aspect of the SCALE-UP classroom is facilitating student interactions with instructors and TAs. We see that this is clearly occurring, especially in small group work, which is one of the most prevalent codes that we observed.

## V. DISCUSSION

Using our new coding scheme, we achieved satisfactory IRR on several codes important to the SCALE-UP class design, such as lecture, feedback, clicker technology, and small group work. However, we were unable to achieve sufficient IRR on other seemingly important codes, like demonstration and student presentation, perhaps due to their unexpected infrequent occurrences.

In SCALE-UP, lecturing should be minimal and instructors should act as facilitators. We found instructors spending nearly half of the intervals lecturing or providing feedback. As expected, we found a reduced amount of lecture across both universities and that instructors used the majority of the SCALE-UP class time to employ small group work. The design of the classroom also helps the instructor and TAs move around the tables to interact with groups and answer their questions promptly. We observed 62% of all intervals in which small group work was observed had a corresponding instructor interaction.

A common approach one might expect to see in SCALE-UP classes is Peer Instruction through the use of clicker technology. Although we observed the use of clickers and clicker questions on average in between 11% and 21% of all observed intervals, we did not observe individual clicker use

transitioning to small group clicker use. Instead, nearly all clicker use coincided with small group work, a finding similar to Turpen's findings [9].

We found that at University A and University B, for lecture, feedback, clicker, and small group work codes, the average percentage of intervals in which the codes were observed were similar for each code. The distributions of each code within each institution were somewhat similar as well. However, clicker use varied largely at University A whereas at University B most professors were focused around the average. For small group work, University B had a fairly wide distribution whereas at University A, aside from two outliers, the instructors averaged a large amount of small group work. Thus, our project must continue to attend to both instructor-level and institution-level teaching patterns.

## VI. CONCLUSION AND FUTURE WORK

As part of a project to characterize how studio classes are taught across the country and to determine which practices correlate to high learning gains, as measured by conceptual tests and attitudinal surveys, we have begun to observe studio classes using a modified TDOP coding scheme. Our initial observations indicate that our coding scheme is sufficient for measuring several important aspects of SCALE-UP classes. However, there are some aspects in which sufficient IRR has not yet been achieved. More work is needed for those codes and training the observers on them.

This is only the beginning of the data collection for this project. We plan to use observations like these, combined with instructor and student interviews, at numerous other institutions to achieve our goals of characterizing studio classes and determining best practices. Future work will go into further improving the coding scheme, performing more observations and interviews, and analyzing which observed practices correlate to high student learning gains.

## ACKNOWLEDGEMENTS

We would like to acknowledge Jon Gaffney for his valuable conversations about this work. This work was supported by the National Science Foundation (Grant Nos. DUE 1347510, 1347515 and 1347527).

- 
- [1] R. J. Beichner *et al.*, *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007).
  - [2] K. Cummings *et al.*, *Am. J. of Phys.*, **67**, S1 (1999).
  - [3] J. M. Saul and R. Beichner, in *Proceedings of the Physics Education Research Conference*, (Rochester, New York, 2001).
  - [4] K. T. Foote, *Phys. Rev. PER* **12**, 1 (2016).
  - [5] J. Ridenour *et al.*, in *Proceedings of the Physics Education Research Conference*, (Philadelphia, PA, 2012).
  - [6] K. T. Foote *et al.*, *Intern. J. STEM Ed.* **1**, 1 (2014).
  - [7] M. T. Hora and J.J. Ferrare, *J. Learn. Sci.* **22**, 2 (2013).
  - [8] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology* (Thousand Oaks, California, 2004).
  - [9] C. Turpen and N.D. Finkelstein, *Phys. Rev. Spec. Top.-PER* **5**, 2 (2009).