

RESEARCH

Open Access



Characterizing science graduate teaching assistants' instructional practices in reformed laboratories and tutorials

Tong Wan¹, Ashley A. Geraets², Constance M. Doty¹, Erin K. H. Saitta² and Jacquelyn J. Chini^{1*} 

Abstract

Background: Graduate teaching assistants (GTAs) often lead laboratory and tutorial sections in science, technology, engineering, and mathematics (STEM), especially at large, research-intensive universities. GTAs' performance as instructors can impact student learning experience as well as learning outcomes. In this study, we observed 11 chemistry GTAs and 11 physics GTAs in a research-intensive institution in the southeastern USA. We observed the GTAs over two consecutive semesters in one academic year, resulting in a total of 58 chemistry lab observations and 72 physics combined tutorial and lab observations. We used a classroom observation protocol adapted from the Laboratory Observation Protocol for Undergraduate STEM (LOPUS) to document both GTA and student behaviors. We applied cluster analysis separately to the chemistry lab observations and to the physics combined tutorial and lab observations. The goals of this study are to classify and characterize GTAs' instructional styles in reformed introductory laboratories and tutorials, to explore the relationship between GTA instructional style and student behavior, and to explore the relationship between GTA instructional style and the nature of laboratory activity.

Results: We identified three instructional styles among chemistry GTAs and three different instructional styles among physics GTAs. The characteristics of GTA instructional styles we identified in our samples are different from those previously identified in a study of a traditional general chemistry laboratory. In contrast to the findings in the same prior study, we found a relationship between GTAs' instructional styles and student behaviors: when GTAs use more interactive instructional styles, students appear to be more engaged. In addition, our results suggest that the nature of laboratory activities may influence GTAs' use of instructional styles and student behaviors. Furthermore, we found that new GTAs appear to behave more interactively than experienced GTAs.

Conclusion: GTAs use a variety of instructional styles when teaching in the reformed laboratories and tutorials. Also, compared to traditional laboratory and tutorial sections, reformed sections appear to allow for more interaction between the nature of lab activities, GTA instructional styles, and student behaviors. This implies that high-quality teaching in reformed laboratories and tutorials may improve student learning experiences substantially, which could then lead to increased learning outcomes. Therefore, effective GTA professional development is particularly critical in reformed instructional environments.

Keywords: Graduate teaching assistants, Reformed laboratories and tutorials, Science laboratory and tutorial instruction, Professional development, Undergraduates

* Correspondence: Jacquelyn.Chini@ucf.edu

¹Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL 32816, USA

Full list of author information is available at the end of the article

Background

Student-centered strategies that promote active learning have been shown to improve student performance in undergraduate science, technology, engineering and mathematics (STEM) (Freeman et al., 2014). Compared to “teaching by telling,” active learning interventions result in higher student performance on exams and concept inventories and lower failure rates (Freeman et al., 2014). Definitions of active learning vary. For example, Freeman et al.’s (2014) meta-analysis classified classroom activities with varied intensity and implementation as active learning, ranging from occasional group problem-solving and use of personal response systems to flipped-style studio or workshop courses. The common feature across these activities is that *some* class time is devoted to students developing their own understanding, rather than a full class period of instructor-led “teaching by telling” (Piaget, 1926; Vygotsky, 1978). Laboratories and recitations are ideal environments to implement active learning strategies, as students are tasked with hands-on activities in small groups. The discipline-based education research community has worked at reforming laboratories and tutorials in STEM in order to foster active learning. In particular, a variety of instructional materials have been developed and demonstrated to improve student conceptual understanding and critical thinking skills (Elby et al., 2007; Etkina & Van Heuvelen, 2007; Greenbowe & Hand, 2005; Holmes, Wieman, & Bonn, 2015; McDermott & Shaffer, 1998). In this paper, we refer to reformed laboratories and tutorials as those in which the curricula implemented were developed based on the findings from research on teaching and learning.

The role of GTAs in student learning

Graduate teaching assistants (GTAs) are often given opportunities to lead laboratories and recitation, especially at large, research-intensive universities (Seymour, 2005). Research has shown that GTAs’ teaching skills are positively correlated with student learning outcomes in reformed tutorial and lab sections (Greenbowe & Hand, 2005; Hazari, Key, & Pitre, 2003; Koenig, Endorf, & Braun, 2007; Poock, Burke, Greenbowe, & Hand, 2007; Stang & Roll, 2014; Wheeler, Maeng, Chiu, & Bell, 2017). Koenig and colleagues, for example, showed that Socratic dialogue implemented by physics GTAs in tutorials has the largest effect on students’ conceptual understanding compared to other teaching strategies (Koenig et al., 2007). Stang and Roll (2014) found that the frequency of GTA-initiated interactions is positively correlated with student engagement in physics laboratory and that student engagement is positively correlated with student learning. Greenbowe and Hand (2005) showed that the more effective the GTAs were at implementing the science writing heuristic in the lab, the higher their

students’ score on their chemistry content assessments. Wheeler, Maeng, Chiu, & Bell (2017) demonstrated that professional development (PD) and teaching improve GTAs’ content knowledge, and GTAs’ post-PD content knowledge is positively correlated with students’ end-of-semester content knowledge. Thus, GTAs’ content knowledge and implementation of the intended pedagogy of active learning teaching methods has been shown to impact student learning.

Studies have also demonstrated that GTAs have impact on student learning experience. O’Neal, Wright, Cook, Perorazio, and Purkiss (2007), for instance, showed that GTAs influence student retention in the sciences through making an impact on lab climate, course grades, and students’ knowledge of science careers. In addition, Kendall and Schussler (2012) reported that GTAs are perceived by students as relatable and understanding; however, students also feel that GTAs are uncertain, hesitant, and nervous, which suggests areas for improved GTA PD.

Considering GTAs’ influence on student learning and classroom experiences, it is important to develop effective programs to prepare GTAs for their roles in active learning environments. Reeves et al. (2016) developed a framework for evaluating and researching GTA PD. They proposed three overarching variable categories for consideration: outcome variables, contextual variables (e.g., institutional setting and GTA training design), and moderating variables (e.g., GTA responsiveness to training). The outcome variables include GTA cognition (e.g., knowledge about specific teaching skills), GTA teaching practices (e.g., what teaching strategies they use in their class), and undergraduate learning outcomes (e.g., knowledge and retention). These three outcome variables are considered the foci of GTA PD evaluation because GTA PD can lead to changes in GTA cognition, which can in turn impact GTA teaching practice, and subsequently impact undergraduate learning outcomes. Reeves et al. (2016) also point out that of the three outcome variables, GTA cognition has been investigated most often, and measuring GTA teaching practices and undergraduate student outcomes face logistic challenges.

Characterizing GTA teaching practices

We argue that research on GTA teaching practice is particularly important as GTA teaching practice has a direct impact on undergraduate learning outcomes. A comprehensive study of GTA teaching practice allows us to gain insights into specific instructional strategies that lead to improved undergraduate learning outcomes and helps researchers develop criteria for effective teaching in science laboratories and tutorials.

Although research on GTA instructional practices is not extensive, a common finding among the studies is

that the extent to which GTAs implement active learning strategies vary substantially (Boman, 2013; Duffy & Cooper, 2019; Luft, Kurdziel, Roehrig, & Turner, 2004; Mutambuki & Schwartz, 2018; Rodriques & Bond-Robinson, 2006; West, Paul, Webb, & Potter, 2013; Wilcox, Kasprzyk, & Chini, 2015). Luft et al. (2004), for example, used qualitative methods to study the educational and instructional environments of science GTAs and found that GTAs made use of the “teaching by telling” approach in labs. Rodriques and Bond-Robinson (2006) assessed chemistry GTAs’ teaching quality from the perspectives of faculty and undergraduates; only 30–40% of the GTAs were found to actively promote students’ conceptual understanding. West et al. (2013) measured instructor-student interactions using a real-time instructor observation tool (RIOT) and found a large variation in the interactions between physics GTAs and students. Wilcox et al. (2015) used RIOT to measure GTA actions in physics tutorials and laboratories; their results suggest that GTAs who were paired with a learning assistant tended to explain less and interact with students less. In all of these studies, GTAs received PD in a variety of formats (e.g., teaching seminar, weekly meetings, formative assessment). However, the variations in implementation of active learning strategies suggest that those PD programs have not met the goal of preparing GTAs to effectively engage students in active learning environments.

As Reeves et al. (2016) pointed out, measuring GTA teaching practice is logistically challenging and studies on teaching practices tend to rely on indirect measurements, such as surveys (e.g., Rodriques & Bond-Robinson, 2006) and interviews (e.g., Luft et al., 2004). Only a few (e.g., Duffy & Cooper, 2019; West et al., 2013 and Wilcox et al., 2015) used direct observations to measure GTA-student interactions. However, these studies do not document other typical behaviors (e.g., GTA real-time writing, student performing lab activity, student taking quiz).

We are not aware of any studies that classify and characterize GTA instructional practices in reformed tutorials and laboratories. However, we found one study that characterizes GTA teaching practices in a traditional general chemistry laboratory (Velasco et al., 2016). In that study, Velasco et al. (2016) developed an observation protocol, the Laboratory Observation Protocol for Undergraduate STEM (LOPUS). LOPUS captures both GTA and student behaviors in 2-min intervals without using previously agreed upon criteria for evaluating teaching quality. In addition, LOPUS documents the initiator of verbal interactions between students and GTAs, as well as the nature of the content of those verbal interactions. Velasco et al. (2016) identified four instructional styles among the GTAs (who implemented

the same curriculum): the waiters, the busy bees, the observers, and the guides on-the-side. However, across all four instructional styles, GTAs initiated fewer verbal interactions than students did. Additionally, student behaviors were found to be independent of GTAs’ instructional styles. Lastly, the nature of verbal interactions was not associated with GTAs’ instructional styles, but were related to the nature of laboratory activities. Since rich observational data has only been collected for the traditional chemistry laboratory environment, more research is needed to explore whether these trends extend to reformed laboratories and tutorials and/or across disciplines.

Our study is intended to bridge the gap in the literature by examining GTAs’ instructional practices in reformed science laboratories and tutorials. We use classroom observations to directly measure GTAs’ and students’ actions. We observe both chemistry and physics classrooms in order to generalize findings to reformed science laboratories and tutorials.

Conceptual framework

We drew from Cohen and Ball’s instructional capacity framework (Cohen & Ball, 1999) to formulate the research questions and to interpret our findings. From Cohen and Ball’s perspective, neither curriculum alone nor instructors alone should be considered the main source of instruction. Instead, instructional capacity is a function of the *interaction* among instructors, students, and instructional materials. Indeed, instructors’ pedagogical content knowledge (Darling-Hammond, 2006) and beliefs about teaching and learning shape how they interact with students and how they enact curricula. Students’ prior knowledge and past educational experiences influence their responses to instructors’ teaching and their engagement with the instructional activities. The interaction between instructor and students is situated in the instructional activities informed by the instructional materials. Therefore, each of the three elements is essential and required for instruction.

Cohen and Ball (1999) stress that instructors play a unique role in the construction of instructional capacity as “their interpretation of educational materials affects curriculum potential and use, and their understanding of students affects students’ opportunities to learn” (p. 4). Instructors have opportunities to learn about student reasoning with particular concepts, common difficulties students encounter, and strategies to address those difficulties.

Cohen and Ball suggest that educational reform is likely to be more effective if it targets the interaction among instructors, students, and curricula, rather than focusing solely on instructors or only on curricula. Interventions that take into account the relations between

those three elements are more likely to improve instructional capacity. Thus, our research questions and interpretations explore the interaction of instructors, students, and curriculum.

Research questions

The current study is part of a larger project on developing an effective GTA PD program that supports GTA success in active learning environments. As an initial step, we characterize GTAs' instructional practices in reformed tutorials and laboratories. We observed chemistry GTAs in reformed laboratories and physics GTAs in both reformed tutorial and laboratory sections using a protocol adapted from LOPUS. Informed by Cohen and Ball's instructional capacity, not only did we classify and characterize GTAs' instructional styles (RQ1) but we also investigated the relationship between GTAs' instructional styles and students' behaviors (RQ2) as well as the relationship between GTAs' instructional styles and the nature of laboratory activity (RQ3), and the relationship between students' behaviors and the nature of the laboratory activity (RQ4). Specifically, we intend to answer the following questions:

1. How do GTAs' instructional practices vary when teaching in reformed laboratories and tutorials?
2. How do GTAs' instructional styles relate to students' behaviors in reformed laboratories and tutorials?
3. What is the relationship between GTAs' instructional styles and the nature of laboratory activity?
4. What is the relationship between students' behaviors and the nature of laboratory activity?

Additionally, we explored whether GTAs' teaching experience may impact their use of instructional styles. We also compare our findings to those of Velasco et al. (2016) to begin the exploration of variation between GTAs' instructional styles in traditional and reformed laboratory sections.

Methods

Context for investigation

Courses and institutional setting

We collected data in a general chemistry laboratory and an introductory physics sequence in a very large, research-intensive institution in the southeastern United States. The general chemistry lab is a stand-alone one-credit course for science majors designed to foster student inquiry and conceptual discovery. Students are required to either have taken general chemistry II (a lecture course) or to be taking general chemistry II concurrently. The lab course meets for 2 h and 50 min every

week. Typically, there are more than 20 sections per semester, and each section is enrolled by up to 24 students. Each lab starts with a 15-min quiz, then follows a 5E learning cycle approach (Bybee & Landes, 1990). At the beginning, a GTA presents a situation or conducts a demonstration leading up to a key question (engagement). Then, students develop a procedure and collect data to answer the key question (exploration). Toward the end of the lab session, the GTA guides the class in sharing their findings, identifying patterns in their data, and summarizing the scientifically accepted understanding (explanation). After the lab session, students read about the topic they discovered in lab (elaboration) and conduct follow-up exercises (evaluation).

The introductory physics sequence investigated in this study was intended for students who major in life sciences. The first course in the sequence (Physics I) mainly covers topics in mechanics, and the second (Physics II) mainly covers electricity and magnetism, circuits, and optics. Students are required to take Physics I before they take Physics II. Each course has two components: lecture and a combined tutorial and laboratory session ("mini-studio"; Chini & Pond, 2014). The lecture component is taught by a faculty member (four lecture sections taught by individual faculty members are represented in this data, two Physics I lecture sections and two Physics II lecture sections), and each mini-studio section is led by a GTA. During this study, each course had two lecture sections, each with class enrollment of 250–300 students. Each lecture section was accompanied by approximately nine sections of mini-studio, and each mini-studio is enrolled by up to 32 students.

The weekly mini-studio is comprised of a 75-min tutorial based on the University of Maryland Open Source Tutorials (Elby et al., 2007), a 15-min group quiz, and an 80-min lab based on the Investigative Science Learning Environment (ISLE) curriculum (Etkina & Van Heuvelen, 2007). The tutorial curriculum focuses on conceptual understanding and sensemaking, and the laboratory curriculum involves guided-inquiry activities intended to improve students' critical thinking skills. Students work in groups of four on the tutorial and lab. The GTA's role is to facilitate students' sensemaking while they apply physics concepts, and design and conduct experiments.

Table 1 gives a brief description of the observed chemistry laboratory activities. All of the activities are taught through guided-inquiry. However, the activities differ in the extent to which conceptual understanding is the main objective. In addition, some laboratory activities require analytical lab techniques and data analysis. The physics tutorials and laboratories have consistent goals (i.e., the tutorials focus on sensemaking of concepts and the laboratories focus on critical thinking skills). Thus, we only categorize the chemistry laboratory activities.

Table 1 Description of chemistry laboratory activities observed

Nature of laboratory activity	Laboratory topic	Brief description of laboratory activity	Number of observations
Conceptual	Electric solutions	Inquire about electrolytes using a conductivity apparatus and various solutions	11
	Thermochemistry	Inquire about how and why acid strength affects the heat of a neutralization reaction	3
Conceptual/ analytical	Environmental solutions I	Conceptualize a calibration curve and calculate serial dilutions	4
	Environmental solutions II	Perform serial dilutions; conduct spectroscopic analysis on an unknown water sample	15
	Chemical kinetics	Explore reaction rates and reaction orders in order to determine how changing the concentration of reagents affects the reaction rate	11
	Hard water ions	Titration analysis to calculate the concentration of hard water ions in tap water	14

Current GTA training design and implementation

During this study, 12 chemistry GTAs were assigned to the lab course per semester. Only one GTA taught the course prior to the study. Typically, each GTA teaches two sections. The chemistry GTAs participate in three 3-h inquiry-based workshops before each semester, which focus on the inquiry method and provide foundational readings. During the semester, GTAs have 90-min weekly preparation meetings, including at least 30 min of pedagogical instruction on topics like questioning students, navigating the zone of proximal development (Gilbert 2006; Vygotsky, 1978), and leading class discussions. Additionally, all first-year GTAs attend a two-semester teaching seminar that consists of monthly meetings, online discussions based on constructivist teaching and learning principles, and evidence-based teaching. These PD activities are led by a faculty (E.K.H.S.) in the chemistry department.

About 13 GTAs teach the mini-studios associated with the two physics courses each semester. During the semesters of this study, approximately half of the GTAs had prior teaching experience in this course sequence. Each GTA typically teaches three sections. All first-year physics GTAs participate in a semester-long weekly pedagogy seminar, led by a faculty (J.J.C.) in the physics department. The seminar engages GTAs in reflecting on their teaching practices, discussing education literature, and evaluating research-based changes to courses. All of the physics GTAs attend weekly 90-min preparation meetings led by an experienced GTA and/or a postdoctoral scholar (C.M.D. and T.W.). During the meetings, GTAs work in small groups led by a facilitator. When working through the tutorials, GTAs take turns to articulate answers and reasoning, and the whole group critiques these answers and reasoning; when working through the laboratories, GTAs design experiments, brainstorm alternatives, and perform experiments. Moreover, GTAs discuss common student difficulties with particular concepts, common student practices in inquiry-based lab activities, and strategies to facilitate student learning.

In addition, all chemistry and physics GTAs attended a pedagogical training activity in a mixed-reality simulator (Chini, Straub, & Thomas, 2016) at the beginning of the second semester during this study. GTAs each led two 7-min discussions with avatar students, during which they rehearsed two pedagogical skills, cold calling and error framing (Becker et al., 2017). We do not report on the rehearsed skills in this paper, and we do not notice any trend in changes in the instructional styles for GTAs who participated in both semesters (as will be discussed in the “Results” section).

Research participants (GTA characteristics)

We conducted classroom observations over two consecutive semesters in one academic year. In each semester, every GTA was observed about four times (range of three to five). A total of 22 GTAs (11 chemistry and 11 physics) volunteered to participate in the study. These GTAs were a mixture of USA citizens and non-citizens. Ten of the chemistry GTAs and six of the physics GTAs were new (first-year) GTAs. Four of the chemistry GTAs and seven of the physics GTAs participated in both semesters. A total of 58 chemistry lab periods and 72 physics combined tutorial and laboratory periods were observed.

Classroom observations using LOPUS

The observation protocol we used was adapted from LOPUS. Codes for GTA behaviors and student behaviors are shown in Tables 2 and 3, respectively. The codes are divided into three categories: typical instructional behaviors, interactive behaviors, and noninstructive behaviors. Considering that our instructional contexts and research goals differ from the study in which LOPUS was developed, the authors examined the face validity of the protocol. Two of the authors (E.K.H.S. and J.J.C.) led the development of the courses we investigated, and the other three authors (A.A.G., C.M.D., and T.W.) have been either the GTAs of the courses or the facilitators for the weekly preparation meetings. Thus, all the authors have expertise to assess the face validity of an

Table 2 Definitions for codes that describe GTA behaviors

Type of behavior	GTA code	Abbreviated definition
Typical instructional behaviors	Lec	Lecturing to the class or making announcements
	RtW	Real-time writing on the board, doc cam, etc.
	FUp	Providing follow-up or feedback on activity
	D/V	Showing a demonstration or video
	M	Monitoring class or individual groups
Interactive behaviors	PQ	Posing a worksheet- or lab-related question
	1o1-Talk	Talking to individual students or group of students one-on-one
	1o1-TPQ	Posing a question to individual students or group of students
	VF	Providing feedback to student responses
	VM	Verbal monitoring
	TI	Initiating one-on-one interaction with students
Noninstructive behaviors	Adm	Performing administrative tasks
	W	Waiting and generally unavailable to students

observation protocol to measure teaching practices in these instructional settings. After a careful review of the original protocol in addition to discussions after practice observations, we decided to make a few modifications to the LOPUS, as described below.

For the GTA behaviors, we eliminated the code “other” as we did not find it necessary during our observations. We changed “verbal monitoring and positive reinforcement” to “verbal monitoring” because the three observers did not come to an agreement on what should be coded as positive reinforcement. We also added a new code, VF (GTA providing verbal feedback), to measure whether GTAs provide clear feedback to student responses. For student behaviors, we changed “performing laboratory activity” to “working on worksheets or performing laboratory activity” for the physics course since both tutorials (that use worksheets) and laboratory are included. We eliminated the code “listening” as we found it highly correlated with the GTA code “lecturing”; additionally, we eliminated the code “prediction” because students were not asked to make predictions as

a whole class in any of the observed courses (although they may have made predictions when they were working in groups, which could be difficult to observe). Lastly, all three codes in the noninstructive behaviors category (i.e., student leave, wait, and other) were eliminated due to the constraint of real-time coding. That is, our researchers were coding while the class was in progress, and thus it could be challenging to code other students’ behaviors when the GTA was interacting with a particular group of students. We note that, in the original LOPUS study, the labs were video-recorded and the researchers conducted video coding.

The modified protocol maintains key characteristics of LOPUS, which captures the behaviors of GTAs and students as well as verbal interactions between GTAs and students. The codes that were eliminated in our study were not coded frequently in the original study (with average percentages ranging from zero to approximately 15%). Moreover, none of those codes except for the student code “listening” was significantly different between clusters in the original study. Another difference

Table 3 Definitions for codes that describe student behaviors

Type of behavior	Student code	Abbreviated definition
Typical instructional behaviors	Wks/Lab	Working on worksheets or performing lab activity
	TQ	Taking a quiz
Interactive behaviors	SQ	Asking the GTA a worksheet- or lab-related question with entire class listening
	1o1-SQ	Individual students or a group of students asking the GTA a worksheet- or lab-related question
	WC	Engaging in whole class discussion in which students are talking serially
	SP ^a	Giving a presentation
	SI	Initiating one-on-one interaction with the GTA

^aOnly used in chemistry lab because students do not give presentations in physics mini-studios

between the modified and the original protocol is that we did not code the nature of verbal interactions (e.g., concept, analysis) due to the constraint of real-time coding. In the original LOPUS, instructional styles were determined based on all behavioral codes, and the analysis for the nature of verbal interactions were separate from instructional styles. Therefore, we can still characterize GTAs' instructional styles using the modified behavioral codes.

The observations were conducted by three of the authors (T.W., A.A.G., and C.M.D.). During the observations, the GTA was asked to wear a microphone so that the observer(s) could hear the GTA through headphones. Students who were close to the GTA were captured by the microphone. The observer(s) used an online platform with the protocol to code GTA and student behaviors. All behaviors occurring in any 2-min interval were coded. The observations were audio recorded, but the audio data are not discussed in this paper.

The observers conducted practice observations prior to the study. The goals of practice observations were to refine the protocol and to achieve inter-rater reliability (IRR). Each practice observation was conducted by pairs or all three observers. Initially, the observers only observed approximately 30 min, and then gradually increased the duration as the observers gained experience. During the practice observations, the observers coded independently and took notes. The observers then discussed disagreements and resolved inconsistencies. A codebook with detailed definitions and specific examples was developed. All the observers familiarized themselves with the codebook before this study.

Investigating inter-rater reliability

In order to investigate IRR, three chemistry lab sections and three physics mini-studio sections were observed by either pairs of observers or all three observers at the beginning of the study (in the first semester). In addition, all three observers observed one chemistry lab section toward the end of the first semester and another two chemistry lab sections toward the middle of the second semester. Each observer first coded the entire lab or combined tutorial and lab independently. We then calculated IRR for each pair of observers. Then, we discussed disagreements and resolved inconsistencies. We used Cohen's Kappa for the overall IRR (i.e., IRR for all the codes during each lab period) and Gwet's AC1 for individual codes (e.g., Adm). We used Gwet's AC1 for individual codes because Cohen's Kappa can be extremely low even when a high percent agreement is achieved if a behavior is coded very frequently or very infrequently (Wongpakaran, Wongpakaran, Wedding, & Gwet, 2013). For chemistry labs, the three observers

achieved an average pre-discussion Cohen's Kappa of 0.79 ± 0.07 , and an average Gwet's AC1 of 0.88 ± 0.14 . For physics combined tutorial and lab, the three observers achieved an average pre-discussion Cohen's Kappa of 0.76 ± 0.14 , and an average Gwet's AC1 of 0.85 ± 0.22 . These results suggest that the three observers were in excellent agreement (greater than 0.75) according to Fleiss' criteria (Wongpakaran et al., 2013).

Cluster analysis

In order to characterize and classify GTA instructional styles, we conducted a cluster analysis using the observation data in R (RStudio Team, 2016). Cluster analysis is a class of techniques that are used to group a set of data objects such that objects in the same group are more similar to each other than to those in other groups. In our case, the data objects were the observed chemistry lab periods or physics combined tutorials and laboratories. We used individual class periods instead of individual GTAs as the units for analysis since each GTA was observed several times. Specifically, the data we used to determine similarity (or dissimilarity) were the fractions of occurrence for GTA and student behaviors observed. To determine the fraction for each code, we first determined the frequency and then divided by the total number of 2-min intervals. For example, if code SQ occurred in seven 2-min intervals and the session lasted for 140 min (70 2-min intervals), then the fraction for code SQ would be 0.1. Since we intended to explore the ways (i.e., codes) in which the clusters may vary, we scaled the fractions such that the average fraction for each code is zero and the standard deviation is one. This would give equal weight to all the codes and thus allow the follow-up statistical tests to tease out codes that vary significantly between clusters even if they do not occur frequently (e.g., student presentation is not expected to occur frequently). We then conducted a cluster analysis on the observation data.

The specific technique we used is hierarchical agglomerative clustering. Agglomerative clustering works in a bottom-up manner. Initially, each chemistry lab or physics mini-studio is considered as its own cluster. The two clusters that are most similar (i.e., have the smallest distance) are then combined into a bigger cluster. The dissimilarity (i.e., distance) in this case is calculated in Euclidean space using the fractions for all codes. This process is iterated until all the labs (or mini-studios) are combined in a single cluster. We used Ward's method (Ward Jr., 1963), which minimizes the total within-cluster variance. We then determined the optimal number of clusters using the *factoextra* package (Kassambara & Mundt, 2017) in R. The specific methods for determining the optimal number of clusters will be discussed in the "Results" section.

We conducted cluster analyses on the chemistry and the physics courses separately due to two major differences between the two contexts: (1) the physics curriculum involved both tutorial and laboratory, but the chemistry curriculum included laboratory only, and (2) there were differences in instructional expectations between the physics lab and chemistry lab (e.g., chemistry GTAs were expected to hold whole-class discussions, which was not the case in the physics lab). The differences between the two courses were considered significant such that a clustering with all the data may lead to two clusters, each dominated by observations from one context. To confirm, we conducted a cluster analysis with all the data and the elbow method (Syakur, Khotimath, Rochman, & Satoto, 2018) suggested three clusters. The results showed that (1) in cluster 1, 52 out of 69 (75%) sessions are from the physics sequence; (2) in cluster 2, 41 out of 45 (91%) sessions are from the chemistry course; and (3) all of the 16 (100%) sessions in cluster 3 are from the physics sequence. This was consistent with our hypothesis that the differences between the contexts are the major factors that influence the clustering of all the data. Furthermore, for each discipline, data were collected in one instructional context rather than a variety of contexts; therefore, it would be inconclusive whether the observed differences were due to discipline or instructional contexts. Therefore, we analyzed data from the two contexts separately.

Statistical analysis

After the observed class periods were grouped in clusters, we used statistical tests to identify the ways (i.e., the codes) in which the clusters are different. Due to a small sample size, we used the Kruskal-Wallis rank sum test (Kruskal & Wallis, 1952). Effect size was determined using the eta-squared for Kruskal-Wallis test (Tomczak & Tomczak, 2014). Post-hoc treatment was done using Dunn's multiple comparison test (Dunn, 1964) with Holm-Bonferroni corrections with the *dunn.test* package (Dinno, 2017) in R.

Results

GTAs' instructional practices

We determined the average fractions for students' and GTAs' behaviors across all observed chemistry lab periods (see Appendix Fig. 1 and Fig. 2). Three student behaviors occurred in more than one-third of the 2-min intervals: performing laboratory activity (Lab, $M = 0.72$, $SD = 0.11$), asking the GTA questions in a one-on-one setting (1o1-SQ, $M = 0.43$, $SD = 0.13$), and initiating one-on-one interactions (SI, $M = 0.33$, $SD = 0.11$). The other student behaviors did not occur frequently (i.e., the average fraction is less than 0.15). Compared to students' behaviors, GTAs' behaviors seemed more diverse:

eight out of 13 codes had averages above 0.15. The three most prevalent GTA behaviors were talking to a student or a group of students one-on-one (1o1-Talk, $M = 0.58$, $SD = 0.14$), lecturing or making announcements (Lec, $M = 0.36$, $SD = 0.13$), and asking a student or a group of students questions one-on-one (1o1-TPQ, $M = 0.33$, $SD = 0.12$). We noticed that the fraction for questions posed by students (1o1-SQ) was similar to the fraction for questions posed by GTAs (1o1-TPQ) in a one-on-one setting. In addition, the fraction for interactions initiated by GTAs (TI, $M = 0.22$, $SD = 0.13$) was also comparable to the fraction for interactions initiated by students (SI). Thus, the chemistry GTAs and students appear to be equally active in the reformed laboratory.

Student behaviors in combined physics tutorials and laboratories (see Appendix Fig. 3) were observed to be very similar to students in the chemistry labs. The three most prevalent behaviors in physics were the same as those in chemistry labs: working on worksheets or performing laboratory activity (Wks/Lab, $M = 0.87$, $SD = 0.09$), asking the GTA questions in a one-on-one setting (1o1-SQ, $M = 0.50$, $SD = 0.17$), and initiating one-on-one interactions (SI, $M = 0.36$, $SD = 0.12$). The other student behaviors all had average fractions below 0.15.

Eight GTA behaviors in physics tutorials and laboratories had averages above 0.15 (see Appendix Fig. 4). The three most prevalent behaviors were talking to a student or a group of students one-on-one (1o1-Talk, $M = 0.67$, $SD = 0.21$), monitoring (M, $M = 0.25$, $SD = 0.18$), and asking a student or a group of students questions one-on-one (1o1-TPQ, $M = 0.25$, $SD = 0.17$). Thus, more GTA codes occurred frequently in the physics tutorial and laboratory sections than the chemistry labs, which suggest the physics GTAs are making use of a wider variety of instructional behaviors.

Students in physics tutorials and laboratories appeared to be more active than physics GTAs. The average fraction for questions posed by students (1o1-SQ) was twice as many as the average fraction for questions posed by GTAs (1o1-TPQ) in a one-on-one setting. In addition, the average fraction for interactions initiated by students (SI) was also approximately twice as many as the average fraction for interactions initiated by GTAs (TI, $M = 0.20$, $SD = 0.13$).

Chemistry GTAs' instructional styles

Figure 1 shows the hierarchical structure of the observed chemistry lab periods, resulting from an agglomerative hierarchical cluster analysis. To explore the number of clusters that may provide meaningful and practical interpretations of the characteristics for each cluster, we used the elbow method (Syakur et al., 2018) and gap statistic method (Tibshirani, Walther, & Hastie, 2001). The results from the elbow method did not suggest an optimal

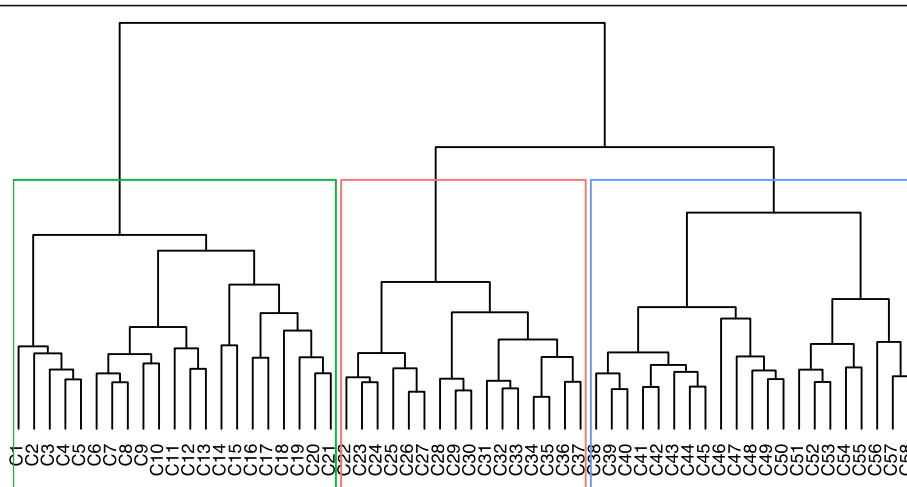


Fig. 1 Dendrogram showing the hierarchical structure of the observed chemistry lab periods

number of clusters, but the results from gap statistic suggested four as the optimal number. We then conducted follow-up statistical tests and interpreted the characteristics for four clusters. However, we found that such grouping led to unnecessarily nuanced distinctions between two of the clusters¹. We then combined the two clusters² and found that the distinctions between the three clusters provided practical interpretations. Therefore, we decided to classify the chemistry GTAs' instructional practices into three different styles (indicated by the boxes in Fig. 1). Using the eta-squared for Kruskal-Wallis test, we found five GTAs' codes and three students' codes that have large effect sizes ($\eta^2 > 0.14$; Maher et al. 2013), and two GTAs' codes and two students' codes that have medium effect sizes ($\eta^2 > 0.06$; Maher et al. 2013) as shown in Figs. 2 and 3, respectively. Below, we characterize the clusters using the seven GTA codes and four student codes with medium or large effect sizes.

The characteristics of chemistry GTAs' instructional practices are shown in Table 4. Each cluster is described by both GTAs' and students' behaviors that occurred significantly differently from other clusters. The complete ranking of clusters for each behavior (resulting from Dunn's test with Holm-Bonferroni corrections) and the median fractions are shown in Table 7 in Appendix.

¹These two clusters are not distinguishable from each other in terms of GTA behaviors. They only differ by the fractions of two student behaviors: one of them has higher average fractions of student asking questions in one-on-one conversations (1o1-SQ) and student initiating interactions (SI) compared to the other.

²The combined cluster is indicated by the blue box on the right in Fig. 3. Before the combination, lab periods C38-C50 were in one cluster and lab periods C51-C58 were in another cluster.

The responders GTAs tended not to initiate interactions with students and provided less verbal feedback. GTAs in cluster A, the responders, initiated the least interactions (TI, $\chi^2(2) = 26.296$, $p_{K-W} < 0.001$) among all clusters (compared to cluster B, $p_{H-B} = 0.007$; compared to cluster C, $p_{H-B} < 0.001$). In addition, GTAs in cluster A provided less verbal feedback (VF, $\chi^2(2) = 15.484$, $p_{K-W} < 0.001$) compared to GTAs in both cluster B ($p_{H-B} < 0.001$) and cluster C ($p_{H-B} = 0.008$). Furthermore, compared to cluster B, GTAs in cluster A posed fewer questions to the whole class (PQ, $\chi^2(2) = 8.807$, $p_{K-W} = 0.01$, $p_{H-B} = 0.005$) and provided less follow-up on activities (FUp, $\chi^2(2) = 8.606$, $p_{K-W} = 0.01$, $p_{H-B} = 0.007$). Correspondingly, students in cluster A asked fewer questions with the whole class listening (SQ, $\chi^2(2) = 11.047$, $p_{K-W} < 0.001$) compared to students in the other two clusters; they also spent more time on the lab activity (Lab, $\chi^2(2) = 9.535$, $p_{K-W} = 0.01$, $p_{H-B} = 0.004$) compared to students in cluster B.

The active lecturers GTAs posed a fair amount of questions to the whole class and tended to write on the board. As mentioned previously, GTAs in cluster B posed more questions to the whole class and provided more follow-up on activity compared to GTAs in cluster A. Additionally, GTAs in cluster B had a higher fraction for real-time writing compared to GTAs in both cluster A (RtW, $\chi^2(2) = 17.52$, $p_{K-W} < 0.001$, $p_{H-B} = 0.01$) and cluster C ($p_{H-B} < 0.001$). In addition, GTAs in cluster B spent less time talking to students one-on-one (1o1-Talk, $\chi^2(2) = 22.666$, $p_{K-W} < 0.001$) compared to both cluster A ($p_{H-B} = 0.01$) and cluster C ($p_{H-B} < 0.001$). Correspondingly, students in cluster B initiated fewer conversations (SI, $\chi^2(2) = 15.230$, $p_{K-W} < 0.001$) and asked fewer questions one-on-one (1o1-SQ, $\chi^2(2) = 17.447$, $p_{K-W} < 0.001$) compared to students in the other

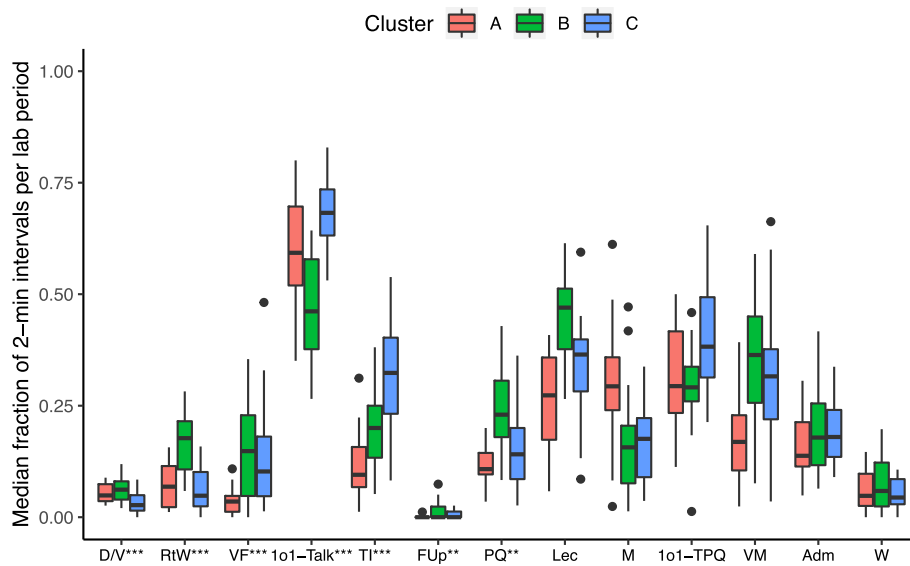


Fig. 2 Median fraction of GTA codes in each cluster. Seven codes are statistically significant among clusters, five of which (D/V, RtW, VF, 1o1-Talk, TI) have large effect sizes and two (FUp, PQ) of which have medium effect sizes. ***Large effect sizes. **Medium effect sizes

two clusters. They also spent less time on the lab activity compared to students in cluster A (as discussed before).

The initiators GTAs consistently initiated interactions with students. The interactions initiated by GTAs in cluster C were the highest among all clusters. GTAs in cluster C also had a higher fraction for talking to students one-on-one than GTAs in cluster B and a higher

fraction for verbal feedback than GTAs in cluster A (as mentioned before). Furthermore, GTAs in cluster C spent less time showing a demonstration or video (D/V, $\chi^2(2) = 12.303$, $p_{K-W} < 0.001$) compared to both cluster A ($p_{H-B} = 0.012$) and cluster B ($p_{H-B} = 0.001$). Correspondingly, students in cluster C initiated more conversations (SI, $p_{H-B} = 0.01$) and asked more questions one-on-one (1o1-SQ, $p_{H-B} = 0.008$) compared to

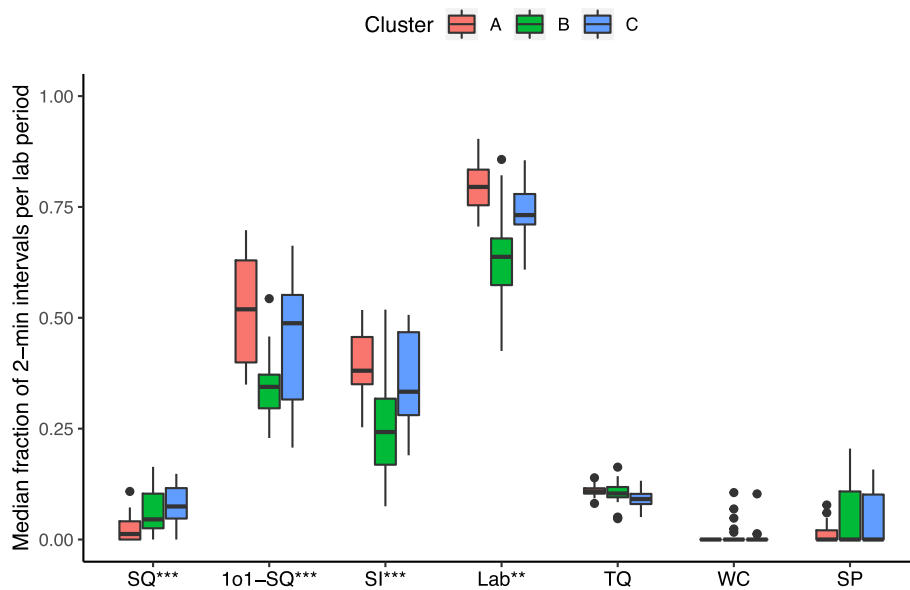


Fig. 3 Median fraction of student codes in each cluster. Four codes are statistically significantly different between clusters, three of which (SQ, 1o1-SQ, SI) have large effect sizes and one has medium effect size (Lab). ***Large effect sizes. **Medium effect sizes

Table 4 Characteristics of chemistry GTAs' instructional practices in each cluster

Cluster	Instructional style	Behaviors that occurred significantly different from other clusters (median fraction)
A (16 sessions)	The responders	GTAs: initiated the fewest one-on-one interactions (0.10) spent more time talking to students one-on-one (0.59) spent less time writing in real time (0.09) spent more time showing a demo or video (0.05) provided less verbal feedback (0.04) provided less follow-up/feedback on activity (0.00) posed fewer questions to whole class (0.11) Students: initiated more one-on-one interactions (0.38) asked more questions in one-on-one interactions (0.52) asked fewer questions with whole class listening (0.01) spent more time performing the lab activity (0.76)
B (21 sessions)	The active lecturer	GTAs: initiated more one-on-one interactions (0.20) spent less time talking to students one-on-one (0.46) spent more time writing in real time (0.12) spent more time showing a demo or video (0.06) provided more verbal feedback (0.18) posed more questions to whole class (0.20) provided more follow-up/feedback on activity (0.00) Students: initiated fewer one-on-one interactions (0.24) asked fewer questions in one-on-one interactions (0.34) spent less time performing the lab activity (0.56) asked more questions with whole class listening (0.05)
C (21 sessions)	The initiators	GTAs: initiated the most one-on-one interactions (0.32) spent more time talking to students one-on-one (0.68) spent less time writing in real time (0.05) spent less time showing a demo or video (0.03) provided more verbal feedback (0.08) Students: initiated more one-on-one interactions (0.33) asked more questions in one-on-one interactions (0.49) asked more questions with whole class listening (0.07)

students in cluster B; they also asked more questions with the whole-class listening (SQ, $p_{H-B} = 0.002$) compared to students in cluster A.

We noticed that both GTAs' and students' behaviors varied across clusters, which suggests that students' behaviors were associated with GTAs' behaviors. Indeed, when GTAs (in cluster C) initiated more interactions and spent more time on one-on-one talk, students asked more questions in both whole-class and one-on-one settings; when GTAs (in cluster A) initiated fewer interactions and provided less verbal feedback, students rarely asked questions in front of the whole class. The observed difference in the amount of questions asked by students in front of the whole class does not seem to be due to a different amount of time spent in the whole class setting. Since the GTAs in cluster A have a similar average fraction for lecturing ($M = 0.27$) as GTAs in the other clusters, students in cluster A, in principle, had opportunities to ask questions in front of the whole class. Instead, the observed difference may be due to the

extent to which GTAs were being interactive (e.g., initiating conversations and asking questions).

We also found that ten out of 11 chemistry GTAs used more than one instructional style, and two GTAs used all three instructional styles, as shown in Fig. 4. The results suggest that the same chemistry GTA may use different instructional styles when leading transformed laboratories. Four of the GTAs participated in both semesters, but we do not notice any trend in changes in instructional styles over two semesters for these GTAs, which may be due to a small sample size.

Physics GTAs' instructional styles

We conducted a cluster analysis for the observed physics tutorials and laboratories with the fractions for all GTAs' and students' codes (Wan, Doty, Geraets, Saitta, & Chini, 2019). We then applied the elbow method and found three clusters (See Fig. 5). We found nine GTA codes and two student codes that have large effect sizes and two GTA codes and two student codes that have medium effect sizes (See Figs. 6 and 7). Compared to the

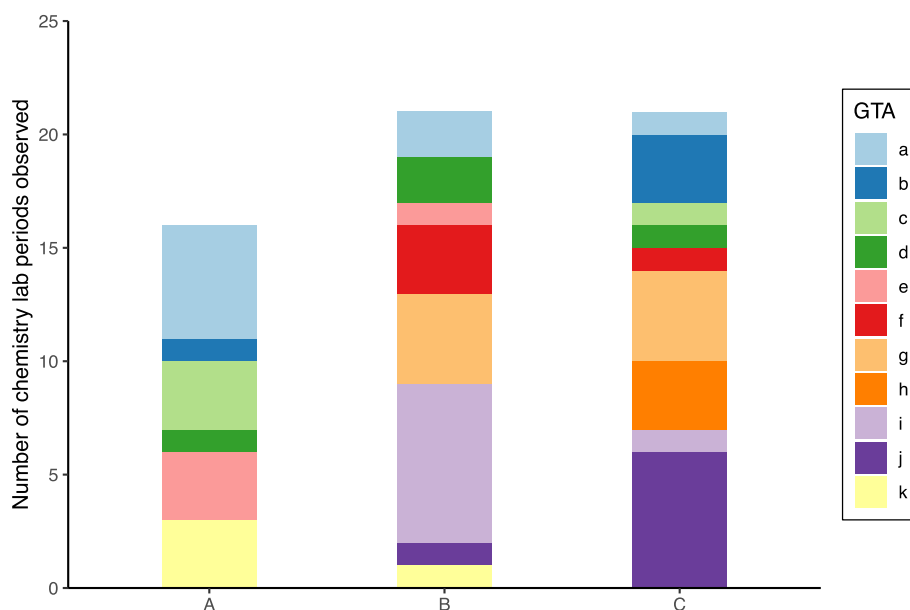


Fig. 4 Number of chemistry lab periods observed led by individual GTAs in each cluster

chemistry GTAs, the variations in physics GTAs' behaviors are more diverse (i.e., nine GTA codes have large effect sizes in physics tutorials and laboratories, while five GTA codes have large effect sizes in chemistry labs). This may be due to that the physics mini-studios involve both tutorial and laboratory while the chemistry context only contains laboratory.

We only use the nine GTA codes and the two student codes with large effect sizes to describe the instructional practices since there are enough codes to characterize three clusters. The characteristics of physics GTAs' instructional practices are shown in Table 5. To highlight

the unique features of each cluster and ease the interpretations, Table 5 only describes behaviors that occurred either more frequently or less frequently than the other *two* clusters (when all three clusters can be ranked, e.g., $A = B > C$), with a few exceptions when only one pair of clusters are distinguishable (e.g., $A > B$, $A = C$, $B = C$). The complete ranking of clusters for each behavior and the median fractions are shown in Table 8 in [Appendix](#).

The group-work facilitators GTAs engaged students in group work. GTAs who were observed teaching the sessions in cluster A had the highest fraction of

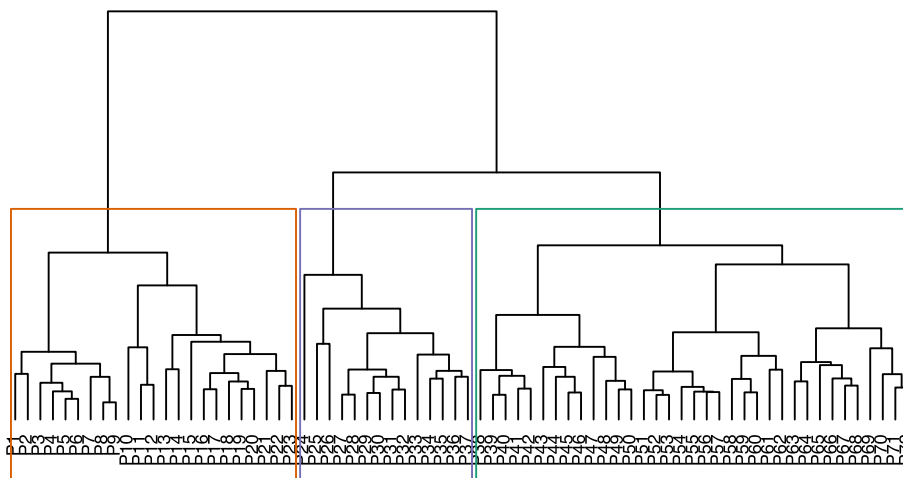


Fig. 5 Dendrogram showing the hierarchical structure of the observed physics mini studios (combined tutorials and laboratories)

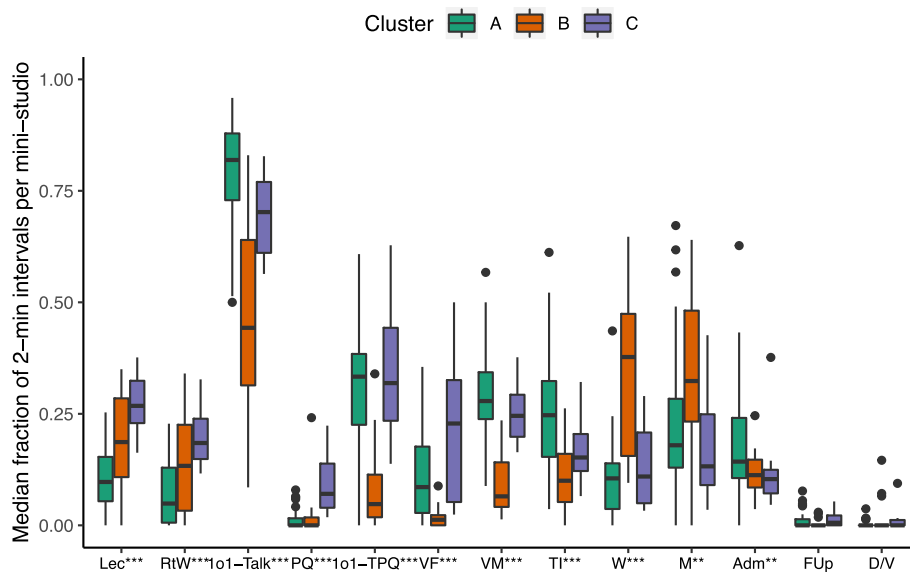


Fig. 6 Median fraction of GTA codes in each cluster in physics mini-studios (combined tutorials and laboratories). Eleven codes are statistically significantly different among clusters, nine of which (Lec, RtW, 1o1-Talk, PQ, 1o1-TPQ, VF, VM, TI, W) have large effect sizes and two of which (M, Adm) have medium effect sizes. ***Large effect sizes. **Medium effect sizes

talking to individual students or groups of students one-on-one (1o1-Talk, $\chi^2(2) = 31.513$, $p_{K-W} < 0.001$) among all clusters (compared to cluster B, $p_{H-B} < 0.001$; compared to cluster C, $p_{H-B} = 0.012$). In addition, GTAs in cluster A initiated more conversations with students (TI, $\chi^2(2) = 20.059$, $p_{K-W} < 0.001$, $p_{H-B} < 0.001$) compared to GTAs in cluster B,

and spent less time writing in real time (RtW, $\chi^2(2) = 16.046$, $p_{K-W} < 0.001$, $p_{H-B} < 0.001$) compared to cluster C. Furthermore, GTAs in cluster A spent the least time lecturing (Lec, $\chi^2(2) = 27.655$, $p_{K-W} < 0.001$) among all clusters (compared to cluster B, $p_{H-B} = 0.003$; compared to cluster C, $p_{H-B} < 0.001$). Correspondingly, students in cluster A asked the

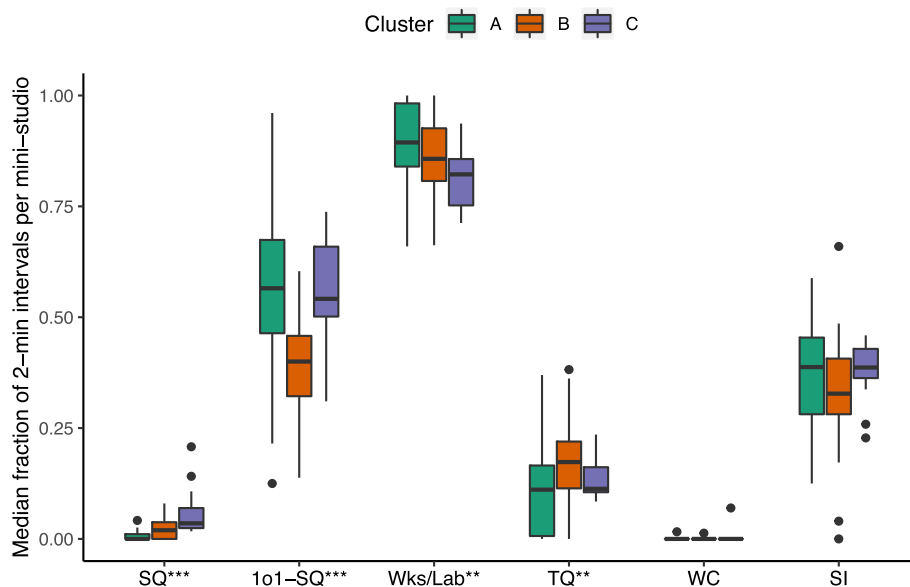


Fig. 7 Median fraction of student code in each cluster in physics mini-studios (combined tutorial and laboratory). Four codes are statistically significantly among different clusters, two of which (SQ, 1o1-SQ) have large effect sizes and two of which (Wks/Lab, TQ) have medium effect sizes. ***Large effect sizes. **Medium effect sizes

Table 5 Characteristics of physics GTAs' instructional practices in each cluster

Cluster	Instructional style	Behaviors that occurred significantly different from other clusters (median fraction)
A (35 sessions)	The group-work facilitators	GTAs: spent most time talking to students one-on-one (0.82) initiated more one-on-one interactions ^a (0.25) spent less time writing in real time ^a (0.05) spent least time lecturing (0.10) Students: asked fewer questions with whole class listening (0.00)
B (23 sessions)	The waiters	GTAs: spent more time waiting (0.38) spent least time talking to students one-on-one (0.44) initiated fewer one-on-one interactions ^a (0.10) asked fewer questions in one-on-one interactions (0.05) provided less verbal feedback (0.01) Students: asked fewer questions in one-on-one interactions (0.40)
C (12 sessions)	The whole-class facilitators	GTAs: asked more questions in front of whole class (0.07) spent more time writing in real time ^a (0.18) had highest fraction of lecturing (0.27) Students: asked more questions with whole class listening (0.04)

^aOnly one pair of clusters are distinguishable; the third cluster is not distinguishable from the other two

least questions with the whole class listening (SQ, $\chi^2(2) = 33.861$, $p_{K-W} < 0.001$) among all clusters.

The waiters GTAs in cluster B tended to wait for students to call on them and interacted with students less frequently. GTAs in cluster B spent significantly more time waiting (W, $\chi^2(2) = 27.576$, $p_{K-W} < 0.001$) compared to both cluster A ($p_{H-B} < 0.001$) and cluster C ($p_{H-B} = 0.001$). In addition, GTAs in cluster B had the lowest fraction of talking to students one-on-one (1o1-Talk) among all clusters. As mentioned previously, GTAs in cluster B had a lower fraction of initiating conversations (TI) compared to cluster A. Moreover, they asked fewer questions in one-on-one interactions (1o1-TPQ, $\chi^2(2) = 32.663$, $p_{K-W} < 0.001$), and provided less verbal monitoring (VM, $\chi^2(2) = 37.736$, $p_{K-W} < 0.001$) and verbal feedback (VF, $\chi^2(2) = 27.776$, $p_{K-W} < 0.001$) compared to both clusters A and C. Correspondingly, students in cluster B asked fewer questions in one-on-one interactions (1o1-SQ, $\chi^2(2) = 18.896$, $p_{K-W} < 0.001$) compared to both cluster A ($p_{H-B} < 0.001$) and cluster C ($p_{H-B} < 0.001$).

The whole-class facilitators GTAs engaged students not only in small groups but also in the whole-class setting and lectured more frequently. Similar to instructional style A, GTAs in cluster C had higher fractions of talking to students one-on-one (1o1-Talk), asking questions in one-on-one conversations (1o1-TPQ), and provided more verbal monitoring

(VM), as well as less waiting (W) compared to cluster B. However, GTAs in cluster C posed more questions (PQ, $\chi^2(2) = 27.227$, $p_{K-W} < 0.001$) than GTAs in both clusters A and B. Furthermore, GTAs in cluster C used the highest fraction of lecturing among all three clusters and had a higher fraction of real-time writing (RtW, $\chi^2(2) = 16.046$, $p_{K-W} < 0.001$, $p_{H-B} < 0.001$) compared to cluster A. Correspondingly, students in cluster C had the highest fraction for asking questions with the whole class listening among all clusters.

Similar to the results from the chemistry laboratories, physics students' behaviors were also found to be associated with GTAs' instructional styles. For clusters A and C in which GTAs were observed to be more engaged (i.e., posing more questions to individuals or groups of students and verbally monitoring students more frequently), individual students asked more questions in small groups as well. Moreover, in cluster C in which GTAs spent more time lecturing and posed more questions to the whole class, students also asked more questions with the entire class listening.

Consistent with the results from the chemistry laboratories, the same physics GTA may use different instructional styles when leading physics mini-studios (combined tutorial and laboratory). The distribution of sessions in each cluster taught by individual physics GTAs can be found in Appendix Figure 5. Seven out of 11 GTAs were found to use

more than one of the instructional styles. Three GTAs used all three instructional styles. Again, we did not notice any trend in changes in instructional styles for GTAs who were observed in both semesters, which may be due to a small sample size.

Relationship between GTA teaching experience and instructional style

To explore the relationship between GTA teaching experience and instructional style, we compared the distributions of new GTAs and experienced GTAs in clusters. We only used data from the physics course because the numbers of experienced and new physics GTAs were comparable while only one out of 11 chemistry GTAs taught the course prior to the study.

Figure 8 shows the distribution of observations for physics GTAs with different teaching experiences across the three clusters. Twenty-nine out of 35 (83%) sessions in cluster A and nine out of 14 (64%) sessions in cluster C were led by new GTAs, while 17 out of 23 (74%) sessions in cluster B were led by experienced GTAs. Since GTAs in both clusters A and C were engaged in facilitating student learning (in group work and/or whole class) while GTAs in cluster B tended to wait, we compared the proportions of new GTAs and experienced GTAs in two groups: cluster B and non-cluster B (i.e., clusters A and C). Due to a small sample size, we used Fisher's exact test (Fisher, 1934). The results suggest that the difference in proportions of new GTAs and

experienced GTAs was significant between the two groups ($p < 0.001$). This implies that sessions led by experienced GTAs were less interactive than sessions led by new GTAs.

Relationship among nature of laboratory activity, GTAs' instructional style, and student behavior

As pointed out in the conceptual framework, the interaction between students and a GTA is situated in the instructional activity, and thus the nature of instructional activity can influence both students' and GTAs' behaviors. Therefore, we examined the relationship between the nature of the laboratory activities and chemistry GTAs' instructional styles. Table 6 shows the number of lab periods for each category of nature of laboratory activity in each cluster. Due to a small number of conceptual lab periods observed, we did not perform a statistical analysis for correlation. Instead, we discuss the relationship qualitatively.

For conceptual lab activities, GTAs were more likely to use the style of the active lecturer or the initiator, and less likely to use the responder. For conceptual/analytical lab activities, the three instructional styles were (approximately) equally likely to be used by the GTAs. As shown in Table 4, the unique characteristics of the responder compared to the other two instructional styles are that GTAs tend not to initiate conversations with students, provide little verbal feedback, and students rarely ask questions in front of the whole class. When the laboratory activity was completely conceptual, GTAs rarely

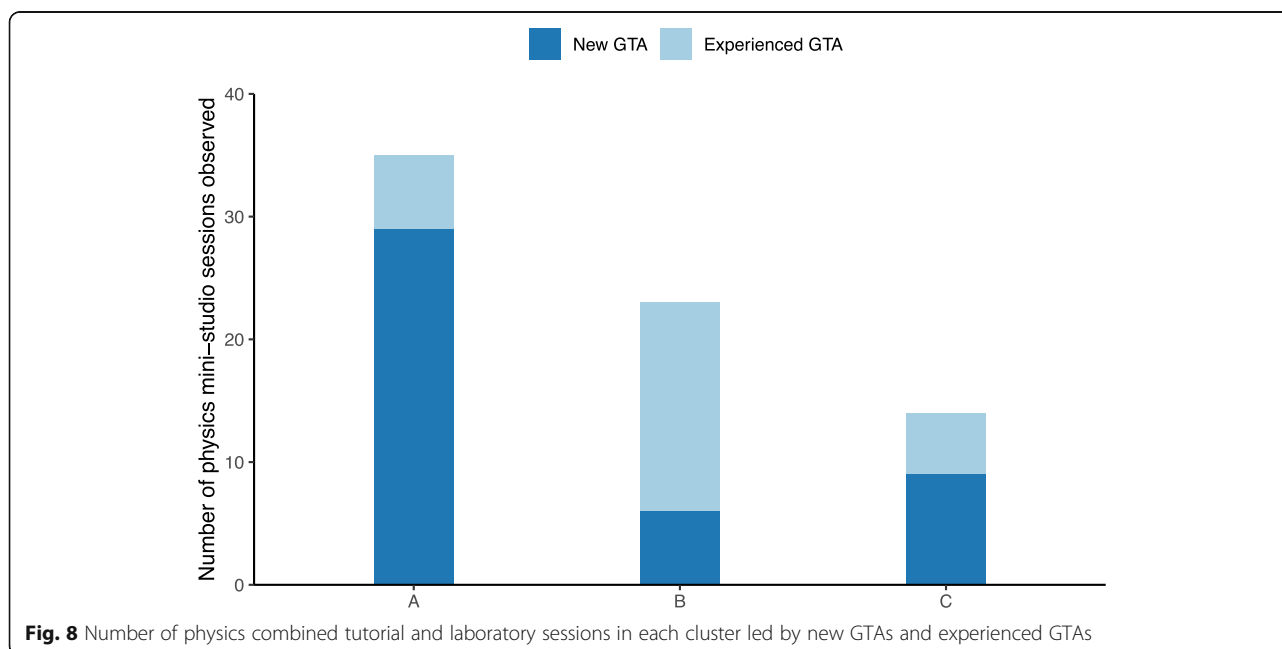


Table 6 Relationship between nature of laboratory activity and chemistry GTAs' instructional style

Nature of laboratory activity	Number of lab periods observed	Instructional style		
		The responders	The active lecturers	The initiators
Conceptual	14	1 (7%)	7 (50%)	6 (43%)
Conceptual/analytical	44	15 (34%)	14 (32%)	15 (34%)

used the style of the responder. Instead, they initiated more interactions and provided more verbal feedback to support students' conceptual understanding. Correspondingly, students raised more questions with the whole class listening. One reason for the observed difference may be that during the experiments that target both concepts and analytical techniques, a portion of the instructional time is spent on the new analytical technique, and therefore less time is allocated for concept development. The results suggest that GTA instructional style and student behaviors may be associated with the nature of laboratory activities.

Discussion

In this study, we used a modified version of LOPUS to characterize science GTAs' instructional practices in reformed laboratories and tutorials. To demonstrate the utility of the modified LOPUS, we applied it in both chemistry and physics courses. Although LOPUS was initially developed in a traditional chemistry laboratory, our study showed that LOPUS can be adapted to characterize instructional practices in reformed laboratories and tutorials as well.

Comparing characteristics of instructional styles in reformed and traditional chemistry laboratories

We discuss the similarities and differences between the characteristics of GTAs' instructional styles in reformed and traditional chemistry laboratory between our study (reformed labs) and the research by Velasco et al. (2016) (traditional labs). We only make qualitative comparisons due to minor differences in the protocols and different data preparations for cluster analysis³.

As mentioned previously, the instructional styles identified in the traditional chemistry lab are the waiters, the busy bees, the observers, and the guides-on-the-side. These instructional styles are characterized by five GTA behaviors: waiting, monitoring, verbal monitoring, GTA initiating interactions, and GTA posing questions in one-on-one conversations.

A major difference in the characteristics of instructional practices between the two instructional contexts is that the instructional practices in the original study were only characterized by GTA behaviors; student behaviors did not vary between clusters. In the reformed laboratory, three student behaviors differ significantly between clusters. Since the reformed laboratory provides opportunities for science practices (e.g., conducting investigations, developing explanatory models), this result indicates that science practices may allow more room for student behaviors to vary.

Overall, the chemistry GTAs in the reformed laboratory made use of more interactive instructional styles compared to the GTAs in the traditional lab. First, the average fraction of waiting in the reformed lab is low (0.06 ± 0.05), and GTAs in different clusters made use of similar amounts (i.e., the difference is not statistically significant) of wait time. Second, both the waiters and the busy bees in the traditional lab rarely provided verbal monitoring (the medians in both clusters were zero), while in the reformed lab verbal monitoring was more prevalent (with an overall average of 0.29 ± 0.16) and the fractions of verbal monitoring were similar across clusters (i.e., the differences between clusters were not statistically significant). Third, the observer in the traditional lab (who is described by monitoring only) had a large fraction (about two-thirds) of monitoring, while in the reformed lab monitoring was less prevalent (with an overall average of 0.20 ± 0.13) and the fractions of monitoring were similar across clusters (i.e., the differences between clusters were not statistically significant). Therefore, the instructional styles in the reformed lab did not seem to resemble the waiters, the busy bees, or the observers. Lastly, the instructional styles (the responders, the active lecturers, and the initiators) in our study appeared to be somewhat similar to the guides-on-the-side in the traditional lab, who asked more questions in one-on-one

³We scaled the data such that the averages are zero and standard deviations are one, while Velasco et al. (2016) did not do so.

⁴Velasco et al. use VP, which is defined as verbal monitoring and positive reinforcement.

⁵1o1-TPQ and VM are not statistically significantly different across the three clusters.

interactions (1o1-TPQ), initiated more interactions (TI), and provided more verbal monitoring (VM⁴). Compared to the guides-on-the-side, all three of the clusters had similar fractions of 1o1-TPQ and VM⁵; the initiators had a similar median fraction of TI as the guides-on-the-side, but the responders and the active lecturers had slightly lower median fractions of TI (see Table 7 in [Appendix](#)).

Thus, all three of the instructional styles identified in our study of reformed chemistry labs share similarities with the most interactive instructional style in the study of traditional chemistry labs. This finding supports the instructional capacity framework since the more interactive curriculum in the reformed labs was associated with more interactive GTA behaviors. In reformed labs, students are often engaged in inquiry activities, and GTAs are trained to facilitate student learning by posing thoughtful questions. Therefore, reformed labs provide more opportunities for GTA-student interactions.

Finally, results from both Velasco et al. (2016) and our study show that the nature of laboratory activities can affect the interaction between GTAs and students, which aligns with the instructional capacity framework. Velasco et al. (2016) showed that the nature of verbal interaction is related to the nature of laboratory activities (e.g., the verbal interaction between GTAs and students centers on concepts when the laboratory activity focuses on concepts). Similarly, we found that GTA instructional style seems to be linked to the nature of laboratory activities: when the activities require in-depth conceptual understanding, GTAs rarely use the style of the responder and are equally likely to use the active lecturer or the initiator, both of which initiate more conversations and provide more verbal feedback; when the activities target not only concepts but also experimental analysis, GTAs are equally likely to use all three instructional styles. This suggests that GTAs use a greater variety of instructional styles when the tasks are more diverse. Student behaviors also appear to be related to the nature of laboratory activities since student behaviors are correlated with GTA instructional style.

Synthesizing findings from reformed chemistry laboratories and physics tutorials and laboratories

We classified and characterized GTAs' instructional styles in both chemistry laboratories and physics tutorials and laboratories. In both instructional contexts, we found that the same GTA may use multiple instructional styles while teaching the same curriculum. These instructional styles differ in a variety of GTA and student behaviors. The results are

consistent with a national study on STEM faculty's teaching practices (Stains et al., 2018) that faculty may use multiple sets of instructional practices across class periods. Stains and colleagues also found that didactic instruction is prevalent among STEM faculty. Our results suggest that instructional practices in tutorials and laboratories are student-centered as students worked on worksheets or laboratory activities most of the time, and lecturing is much less prevalent. However, the extent to which GTAs use questioning techniques to support active learning (rather than directly explaining content) varies in both physics and chemistry contexts. The results suggest that more effective GTA PD is needed to support GTAs in active learning environments.

We noticed that the characteristics of the instructional styles in the chemistry and physics contexts are different. This aligns with our decision of conducting cluster analyses on the two instructional contexts separately. Moreover, chemistry GTAs tend to be more active than physics GTAs since chemistry GTAs spend little time waiting while some of the physics GTAs tend to wait until they are called on by students. Informed by Reeves et al. (2016), we discuss the contextual variables (e.g., GTA training content and structure) and the moderating variables (e.g., GTA responsiveness) in the two GTA PD programs in order to understand the difference in chemistry and physics GTAs' instructional styles. The first-year chemistry GTAs attended a two-semester monthly teaching seminar while the first-year physics GTAs attended a one-semester weekly pedagogy seminar. In addition, the chemistry weekly preparation meetings were led by a faculty member while the physics weekly preparation meetings were led by a graduate student and/or a post-doc. Prior research on GTA "buy-in" has suggested the importance of departmental climate on GTA instructional strategies in reformed instructional environments (Goertzen, Scherr, & Elby, 2009). Our results suggest that GTAs may receive signals about the importance of their adherence to curricular expectations from departmental norms such as positionality of the person in charge of the GTA preparation and length of the GTA pedagogy seminar.

To explore factors that impact GTAs' use of instructional styles, we examined the relationships between GTA instructional styles and student behaviors as well as the nature of laboratory activities. We found that GTAs' instructional styles are associated with student behaviors: when GTAs use more interactive styles, students ask more questions. This finding supports the conceptual framework that

instructional capacity depends on the interaction between instructors and students. However, it contrasts the finding by Velasco et al. (2016) that student behavior is independent of GTA instructional styles in a traditional chemistry laboratory. As argued by Velasco et al. (2016), the interaction between GTA and student behaviors may be limited in the traditional laboratory since the curriculum can influence GTA-student interaction. Indeed, we also found that the nature of laboratory activities appears to be associated with GTA instructional styles and student behaviors: when the focus of the laboratory activity is concepts rather than analytical technique and data analysis, GTAs initiate more interactions and provide more verbal feedback, and students raise more questions in front of the whole class.

Lastly, we examined the difference in instructional styles between new and experienced GTAs using the data from the physics courses since the numbers of new and experienced physics GTAs were comparable (which was not the case in the chemistry course). The results suggest that new GTAs seem more interactive than experienced GTAs. We conjecture that this may be due to lack of incentives for high-quality teaching and increasing research tasks as graduate students progress through graduate school. Furthermore, this may reflect a university and/or departmental culture that research is highly prioritized and that support for teaching is limited (Gardner & Jones, 2011; Nyquist et al., 1999). Another reason may be that GTAs' behaviors are influenced by student expectations. New GTAs have limited experience interacting with students and may be more receptive to PD, while experienced GTAs may have learned student expectations, which can influence their behaviors. This speculation is consistent with findings from prior research. For example, Wheeler, Maeng, and Whitworth (2017) found that some GTAs whose teaching beliefs shifted toward "TA as facilitator" during TA PD reverted back to "TA as disseminator" after teaching. Similarly, Wilcox, Yang, and Chini (2016) found that GTAs' behaviors are influenced by their perceptions of student expectations. Our results suggest that when a GTA makes use of more interactive instructional styles, students are more active as they asked more questions and initiated more conversations. These findings together appear to support the two-way interplay between GTA and student behaviors: instructors who behave interactively can have a positive impact on student behaviors, but some instructors may become less interactive in response to student expectations. Therefore, we suggest that future research further explores how GTAs' behaviors can influence students' behaviors and in return, how students' behaviors further impact GTAs' behaviors. This also suggests that it may be useful for GTA

PD to emphasize GTA agency in shifting student expectations and behaviors.

Limitations

An important limitation of this study may be that all three observers were facilitators for the weekly GTA preparation meetings, which may have had an impact on GTA behaviors. In addition, only GTAs were wearing microphones during observations; therefore, only voices from GTAs and students close to GTAs were captured. Moreover, student group dynamics may also interact with GTA behaviors, but they were not documented in this study. Furthermore, we only investigated one curriculum per discipline at one institutional setting; as pointed out by the Reeves et al. (2016) framework, it is necessary to consider contextual (e.g., institutional setting, GTA training design, GTA characteristics) and moderating variables (e.g., GTA characteristics and implementation variables) to investigate the extent to which successful PD programs transfer across institutions and departments. Future work should explore GTA instructional styles in a variety of curricula within multiple disciplines to more thoroughly address questions about disciplinary norms and the relationship between GTA behaviors, student behaviors, and curricula. Since the authors investigated the face validity of the presented version of the modified LOPUS based on their expertise in the instructional areas investigated here, additional investigation into the face validity of the protocol will be needed before it is used in other instructional settings. Lastly, we did not report the nature of verbal interaction between GTA and students (e.g., underlying scientific principles, data analysis and calculations), which was documented in Velasco et al. (2016). We plan to analyze the nature of verbal interactions in the future.

Conclusion

In this study, we characterized instructional practices in reformed laboratories and tutorials in two science disciplines. We found that science GTAs use a variety of instructional styles when teaching in the reformed laboratories and tutorials. Compared to traditional laboratories, reformed laboratories appear to allow for more interaction between the nature of lab activities, GTA instructional styles, and student behaviors. This implies that high-quality teaching in reformed laboratories and tutorials may improve student learning experiences substantially, which could then lead to increased learning outcomes. Therefore, effective GTA PD is particularly critical in reformed instructional environments.

Appendix

Table 7 Median fractions in each cluster of chemistry labs. The ranking of clusters for each code is determined by Dunn's test with Holm-Bonferroni corrections

	Code	Ranking	Cluster A The responders	Cluster B The active lecturers	Cluster C The initiators
GTA behaviors	TI	C>B>A	0.10	0.20	0.32
	1o1-Talk	A=C>B	0.59	0.46	0.68
	RtW	B>A=C	0.09	0.12	0.05
	D/V	A=B>C	0.05	0.06	0.03
	VF	B=C>A	0.04	0.18	0.08
	FUp	B>A	0.00	0.00	0.00
	PQ	B>A	0.11	0.20	0.13
Student behaviors	SI	A=C>B	0.38	0.24	0.33
	1o1-SQ	A=C>B	0.52	0.34	0.49
	Lab	A>B	0.76	0.56	0.72
	SQ	B=C>A	0.01	0.05	0.07

Table 8 Median fractions in each cluster of physics mini-studios. The ranking of clusters for each code is determined by Dunn's test with Holm-Bonferroni corrections

	Code	Ranking	Cluster A The group-work facilitators	Cluster B The waiters	Cluster C The whole-class facilitators
GTA behaviors	1o1-Talk	A>C>B	0.82	0.44	0.70
	TI	A>B	0.25	0.10	0.15
	1o1-TPQ	A=C>B	0.33	0.05	0.32
	VM	A=C>B	0.28	0.06	0.25
	VF	A=C>B	0.09	0.01	0.23
	PQ	C>A=B	0.00	0.00	0.07
	RtW	C>A	0.05	0.13	0.18
	Lec	C>B>A	0.10	0.19	0.27
	W	B>A=C	0.11	0.38	0.11
	1o1-SQ	A=C>B	0.57	0.40	0.54
Student behaviors	SQ	C>B>A	0.00	0.02	0.04

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40594-020-00229-0>.

Additional file 1: Figure 1. Average fraction of 2-min intervals per chemistry lab period for each student code. Error bars represent standard deviations

Additional file 2: Figure 2. Average fraction of 2-min intervals per chemistry lab period for each GTA code. Error bars represent standard deviations

Additional file 3: Figure 3. Average fraction of 2-min intervals per physics mini-studio (combined tutorial and laboratory) for each student code. Error bars represent standard deviations

Additional file 4: Figure 4. Average fraction of 2-min intervals per physics mini-studio (combined tutorial and laboratory) for each GTA code. Error bars represent standard deviations

Additional file 5: Figure 5. Number of physics combined tutorial and laboratory sessions in each cluster led by individual GTAs

Abbreviations

STEM: Science, technology, engineering, and mathematics; GTA: Graduate teaching assistants; LOPUS: Laboratory Observation Protocol for Undergraduate STEM; PD: Professional development; ISLE: Investigative science learning environment; IRR: Inter-rater reliability; M: Mean; SD: Standard deviation; Lec: Lecture; D/V: Demonstration/video; RtW: Real-time writing; VF: Verbal feedback; 1o1-Talk: One-on-one talk; TI: TA initiating; FUp: Follow-up; PQ: Posing questions; 1o1-TPQ: One-on-one TA posing questions; VF: Verbal feedback; VM: Verbal monitoring; W: Wait; M: Monitoring; Adm: Administration; 1o1-SQ: One-on-one student question; SI: Student initiating; SQ: Student question; Wks/Lab: Worksheet/laboratory; TQ: Taking quiz

Acknowledgements

The authors acknowledge Andrew Cheshire for support in organizing the observation data for analysis.

Authors' contributions

JJC and EKHS conceived the study design. TW, AAG, and CMD conducted the observations. TW conducted the analysis, and TW, JJC, and EKHS interpreted the findings. All the authors read and approved the final manuscript.

Funding

This work is supported in part by the National Science Foundation award #1725554.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available due to human research protections but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL 32816, USA. ²Department of Chemistry, University of Central Florida, 4111 Libra Dr., Orlando, FL 32816, USA.

Received: 18 November 2019 Accepted: 5 June 2020

Published online: 29 June 2020

References

- Becker, E. A., Easlon, E. J., Potter, S. C., Guzman-Alvarez, A., Spear, J. M., Facciotti, M. T., Igo, M. M., Singer, M., & Pagliarulo, C. (2017). The effects of practice-based training on graduate teaching assistants' classroom practices. *CBE Life Sciences Education*, 16(4). <https://doi.org/10.1187/cbe.16-05-0162>.
- Boman, J. S. (2013). Graduate student teaching development: Evaluating the effectiveness of training in relation to graduate student characteristics. *The Canadian Journal of Higher Education*, 43(1), 100–114.
- Bybee, R. W., & Landes, N. M. (1990). Science for life & living: An elementary school science program from biological sciences curriculum study. *The American Biology Teacher*, 52(2), 92–98. <https://doi.org/10.2307/4449042>.
- Chini, J. J., & Pond, J. W. T. (2014). Comparing traditional and studio courses through FCI gains and losses. In *Physics education research conference proceedings* (pp. 51–54).
- Chini, J. J., Straub, C. L., & Thomas, K. H. (2016). Learning from avatars: Learning assistants practice physics pedagogy in a classroom simulator. *Physical Review Physics Education Research*, 12, 010117. <https://doi.org/10.1103/PhysRevPhysEducRes.12.010117>.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement; CPRE research report series RR-43; consortium for policy research in Education*: University of Pennsylvania, Graduate School of Education.
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57, 300. <https://doi.org/10.1177/0022487105285962>.
- Dinno, A. (2017). *DunnTest: Dunn's test of multiple comparisons using rank sums R* package version 1.3.5. <https://CRAN.R-project.org/package=dunn.test>.
- Duffy, E. M., & Cooper, M. M. (2019). Assessing TA buy-in to expectations and alignment of actual teaching practices in a transformed general chemistry laboratory course. *Chemical Education Research and Practice*, 21, 189–208. <https://doi.org/10.1039/C9RP00088G>.
- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6, 241–252.
- Elby, A., Scherr, R. E., McCaskey, T., Hodges, R., Redish, E. F., Hammer, D., & Bing, T. (2007). *Open source tutorials in physics Sensemaking: Suite I* https://www.physport.org/curricula/MD_OST/.
- Etkina, E., & Van Heuvelen, A. (2007). Investigative science learning environment—A science process approach to learning physics. In E. F. Redish & P. Cooney (Eds.), *PER-based reforms in calculus-based physics* (Vol. 1, pp. 1–48). College Park: American Association of Physics Teachers.
- Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Edinburgh: Oliver & Boyd.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 8410.
- Gardner, G. E., & Jones, M. G. (2011). Pedagogical preparation of science graduate teaching assistant: Challenges and implications. *Science Education*, 20, 31–41.
- Gilbert, J. K. (2006). On the nature of "context" in chemical education. *International Journal of Science Education*, 28(9), 957–976. <https://doi.org/10.1080/09500690600702470>.
- Goertzen, R. M., Scherr, R. E., & Elby, A. (2009). Accounting for tutorial teaching assistants' buy-in to reform instruction. *Physical Review Physics Education Research*, 5, 020109. <https://doi.org/10.1103/PhysRevSTPER.5.020109>.
- Greenbowe, T. J., & Hand, B. (2005). Introduction to the science writing heuristic. In N. J. Pienta, Cooper, M. M. & Greenbowe, T. J. (Eds.), *Chemists' guide to effective teaching* (1st ed.). Prentice-Hall: Upper Saddle River, NJ.
- Hazari, Z., Key, A. W., & Pitre, J. (2003). Interactive and affective behaviors of teaching assistants in first year physics laboratory. *Electronic Journal of Science Education*, 7(3), 1–38.
- Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 11199.
- Kassambara, A., & Mundt, F. (2017). *Factoextra: Extract and visualize the results of multivariate data analyses R* package version 1.0.5. <https://CRAN.R-project.org/package=factoextra>.
- Kendall, K. D., & Schussler, E. E. (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE Life Sciences Education*, 11(2), 187–199. <https://doi.org/10.1187/cbe.11-10-0091>.
- Koenig, K. M., Endorf, R. J., & Braun, G. A. (2007). Effectiveness of different tutorial recitation teaching methods and its implications for TA training. *Physical Review Physics Education Research*, 3, 010104. <https://doi.org/10.1103/PhysRevSTPER.3.010104>.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621.
- Luft, J. A., Kurdziel, J. P., Roehrig, G. H., & Turner, J. (2004). Growing a garden without water: Graduate teaching assistants in introductory science

- laboratories at a doctoral/research university. *Journal of Research in Science Teaching*, 41(3), 211–233. <https://doi.org/10.1002/tea.20004>.
- Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE Life Sciences Education*, 12(3). <https://doi.org/10.1187/cbe.13-04-0082>.
- McDermott, L. C., & Shaffer, P. S. (1998). *Tutorials in introductory physics*. Englewood Cliffs: Prentice-Hall.
- Mutambuki, J. M., & Schwartz, R. (2018). We don't get any training: The impact of a professional development model on teaching practices of chemistry and biology graduate teaching assistants. *Chemical Education Research and Practice*, 19(1), 106–121. <https://doi.org/10.1039/C7RP00133A>.
- Nyquist, J. D., Manning, L., Wulff, D. H., Austin, A. E., Sprague, J., Fraser, P. K., Calcagno, C., & Woodford, B. (1999). On the road to becoming a professor: The graduate student experience. *Change*, 31, 18–27. <https://doi.org/10.1080/00091389909602686>.
- O'Neal, C., Wright, M., Cook, C., Perorazio, T., & Purkiss, J. (2007). The impact of teaching assistants on student retention in the sciences: Lessons for TA training. *Journal of College Science Teaching*, 36(5), 24–29.
- Piaget, J. (1926). *The language and thought of the child*. New York: Harcourt Brace.
- Poock, J. R., Burke, K. A., Greenbowe, T. J., & Hand, B. M. (2007). Using the science writing heuristic in the general chemistry laboratory to improve students' academic performance. *Journal of Chemical Education*, 84, 1371–1379. <https://doi.org/10.1021/ed084p1371>.
- Reeves, T. D., Marbach-Ad, G., Miller, K. R., Ridgway, J., Gardner, G. E., Schussler, E. E., & Wischusen, E. W. (2016). A conceptual framework for graduate teaching assistant professional development evaluation and research. *CBE Life Sciences Education*, 15(2), es2. <https://doi.org/10.1187/cbe.15-10-0225>.
- Rodrigues, R. A. B., & Bond-Robinson, J. (2006). Comparing faculty and student perspectives of graduate teaching assistants' teaching. *Journal of Chemical Education*, 83(2), 305. <https://doi.org/10.1021/ed083p305>.
- RStudio Team. (2016). *RStudio: Integrated development for R*. Boston: RStudio, Inc. <http://www.rstudio.com/>.
- Seymour, E. (2005). *Partners in innovation: Teaching assistants in college science courses*. Rowman & Littlefield, Boulder.
- Stains, M., et al. (2018). Anatomy of STEM teaching in north American universities. *Science*, 359, 8383. <https://doi.org/10.1126/science.aap8892>.
- Stang, J. B., & Roll, I. (2014). Interactions between teaching assistants and students boost engagement in physics labs. *Physical Review Physics Education Research*, 10, 020117. <https://doi.org/10.1103/PhysRevSTPER.10.020117>.
- Syakur, M. A., Khotimath, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, 336, 012017.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, B*, 63(Part 2), 411–423.
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 1(21), 19–25.
- Velasco, J. B., Knedeisen, A., Xue, D., Vickrey, T. L., Abebe, M., & Stains, M. (2016). Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate STEM. *Journal of Chemical Education*, 93(7). <https://doi.org/10.1021/acs.jchemed.6b00062>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wan, T., Doty, C. M., Geraets, A. A., Saitta, E. K. H., & Chini, J. J. (2019). Characterizing graduate teaching assistants' teaching practices in physics "mini-studios". In *Physics education research conference proceedings*.
- Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236.
- West, E. A., Paul, C. A., Webb, D., & Potter, W. H. (2013). Variation of instructor-student interactions in an introductory interactive physics course. *Physical Review Physics Education Research*, 9, 010109. <https://doi.org/10.1103/PhysRevSTPER.9.010109>.
- Wheeler, L. B., Maeng, J. L., Chiu, J. L., & Bell, R. L. (2017). Do teaching assistants matter? Investigating relationships between teaching assistants and student outcomes in undergraduate science laboratory classes. *Journal of Research in Science Teaching*, 54(4), 463–492. <https://doi.org/10.1002/tea.21373>.
- Wheeler, L. B., Maeng, J. L., & Whitworth, B. A. (2017). Characterizing teaching assistants' knowledge and beliefs following professional development activities within an inquiry-based general chemistry context. *Journal of Chemical Education*, 94(1). <https://doi.org/10.1021/acs.jchemed.6b00373>.
- Wilcox, M., Kasprzyk, C. C., & Chini, J. J. (2015). Observing teaching assistant differences in tutorials and inquiry-based labs. In *Physics education research conference proceedings* (pp. 371–374).
- Wilcox, M., Yang, Y., & Chini, J. J. (2016). Quicker method for assessing influences on teaching assistant buy-in and practices in reformed courses. *Physical Review Physics Education Research*, 12, 020123. <https://doi.org/10.1103/PhysRevPhysEducRes.12.020123>.
- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)